The Formats of Position Representations in Vision

by

Feitong Yang

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March, 2018

Abstract

Does perception describe locations as distances from orthogonal axes, in a Cartesian format? Or does it employ polar descriptions, as distances from the origin with angular bearings? Are these representations even distinguishable? After all, they are isomorphic: any representation in one format can be translated to the other. To answer these questions, in this study we investigated what coordinate systems are used to describe positions in our visual processing, and when this occurs.

First, we examined noisy responses in simple localization tasks. Variances of responses were better explained by a model that drew responses from noisy polar representations compared to the Cartesian model, except for localization by saccade, where the Cartesian model was better. This contrast demonstrates that Cartesian and polar representations can be distinguished, and that different systems can rely on different representational formats.

We further examined different coordinate systems by considering it as an explanation for Vernier acuity performance. Across a wide range of conditions, including Vernier stimuli placed on oblique axes, we found significant biases in reporting alignment for stimuli that were misaligned towards the circumference of the circle defined by fixation —stimuli misaligned in Cartesian terms, but aligned (i.e. linearly related) in polar terms.

Finally, we studied computational complexity in recognition algorithms. As inputs to template matching routines, different formats differ in terms of which templates are described linearly vs. non-linearly. We reasoned that linear templates should be more easily matched than non-linear ones. We therefore replicated and extended experiments on the identification of overlapping Glass patterns. Participants recognized circular and radial patterns more easily than line patterns, consistent with the polar coordinate hypothesis that circle and radial patterns have linear descriptions.

Discussing representations, Marr (1982) stated “how information is represented can greatly affect how easy it is to do different things with it.” Our results suggest that the format of visual spatial representation makes it easy to do things that are linear in polar terms.

Committee Members:

Jonathan Flombaum (Advisor)

Howard Egeth (Second Reader)

Michael McCloskey

Steven Gross

Jason Fischer

Alternate Members:

Michael Williams

Shreesh Mysore

Acknowledgements

Understanding how the mind works is an arduous journey. It was an unexpected one for me, but full of intellectual excitement and joy. I want to take this opportunity to express my gratitude to the people who have walked with me and helped me along the way.

First of all, I owe a great deal to Dr. Jonathan Flombaum, who brought me into the world of computational functionalism. I enjoyed Jon's perspectives on studying computational processes of human behavior right from the start, but it took me years of training, including courses and discussions, to truly understand the essence of this approach and to be able to adopt this perspective to understand human cognition and behavior, both technically and philosophically. This slow awakening was accompanied by frictions and confusions in my collaboration with Jon. I deeply appreciate his patience and tolerance during our five-year collaboration. I am also grateful for the freedom that Jon granted to me so that I could explore many different research fields and extend my knowledge in mathematical modeling. In addition to science, I have also learned important skills from Jon about how to present ideas and arguments in various formats, such as posters, talks, and sometimes writing. Jon has taught me how to be a real scientist and has cleaned out my previous superficial understanding of doing science. He has also given me a good taste of science. As a result, I feel confident in continuing my research both within and outside of academia in the future.

This thesis would not have been possible without working with Jon and Dr. Michael McCloskey. Mike's book, *Visual Reflections*, provided a theoretical starting point for this work. The main computational problems in my thesis come from this book. A continuing collaboration with Mike on the MDK project inspired me to develop pilot versions of experiments in my thesis research. I am also fortunate to have worked with Celia Litovsky in testing sessions for more than two years, during which I learned a great deal from her about neuropsychological research. I feel lucky and grateful to MDK and his wife Nina for the past two years to explore the mystery of MDK's pointing, saccades, reading, and many other cognitive abilities. It is my first time working with a neurological case. The observed behavioral phenomena are amazing and thought-provoking, and have driven me to continuously think about how our minds process space.

I am grateful to Dr. Howard Egeth and Dr. Jason Fischer for being on my thesis committee and supervising my thesis work from the beginning. Howard and Jason provided important guidance and suggestions to help me develop my thesis. I also want to thank Drs. Steven Gross, Michael Williams, and Shreesh Mysore for being on my thesis committee and providing helpful comments and suggestions on my work.

One major advantage of studying in the Department of Psychological and Brain Sciences is that one has good opportunities to talk to any faculty member. Hence, I want to thank many such members for their time and intellectual discussions. For example, I learned about neural plasticity from Dr. Marina Bedny, Bayesian modeling from Dr. Colin Wilson. Furthermore, those conversations I had with Dr. Justin Halberda have made great memories of my time at graduate school, from which I learned not only great thoughts in cognitive science, but also the importance of thinking logically and articulating concisely. I often regretted not being able to have more in-depth discussions with other faculty members in our department. I believe each discussion exerted its impact on my current and future work.

My peers are the most important source of support for me to carry on my daily life in research. I feel highly fortunate to have a supportive cohort. Among others, I mostly want to thank Dr. Zheng Ma. As a senior member in the lab, she consistently helped me, supported me, and encouraged me. We also had many insightful discussions that were very helpful for me to push forward in my research. I also want to thank Jenny Wang for her intellectual discussions and tasty food that fed both my mind and my stomach. My peers, Tammy Tran, Shipra Kanjlia, and Ela Warnecke, together with Vyash Puliyadi, gave me helpful social support when I was down, anxious, or confused. Friendships with other graduate students and postdocs, such as Gi-Yeul Bae, Ninad Kothari, Jinhong Luo, Mark Schurgin, and many others, are all irreplaceable to me.

Love and accompaniment have given me strength during my journey and have encouraged me to overcome difficulties. I want to specially thank Qian Yu for her love, understanding, care, and support. Meeting Qian has been one of the greatest rewards in my journey, and walking with her beautiful mind makes my journey and life much more enjoyable.

Finally, I have been blessed to be supported by the love of my family. My father, Xuejun Yang, always supports what I study, and always encourages me to pursue a deep understanding of what I care about and study. His unconditional love has taught me important lessons in my life and provided me with the power to love others. I also need to thank my other family members, who take care of my father so that I can enjoy my journey without worry.

Finishing my Ph.D. dissertation is just the beginning. Just like obtaining a black belt in a martial art practice, I believe that I have finally acquired the fundamental skills and that I am now ready to start the real, exciting, and challenging part of my journey towards the understanding of minds, machines, and brains.

Feitong Yang

March 5th, 2018

Baltimore, Maryland

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Perceiving positions, motion, and spatial relationships is quick and easy. People do not think about how they process such information, nor do they try to perceive locations: it just happens. Such "effortlessness" obscures tremendous computational challenges in vision. One problem comes from impoverished input. When the external world is projected on a retina, a three-dimensional space is reduced to a two-dimensional light pattern that subsequently activates a two-dimensional neuronal pattern. Losing depth in this projection process leads to occlusions between objects and leaves incomplete traces of external objects on the retinae. Worse still, losing depth makes visual inference problems underdetermined, as illustrated in Figure 1.1: one retinal image, in principle, corresponds to infinitely many possible external stimuli.

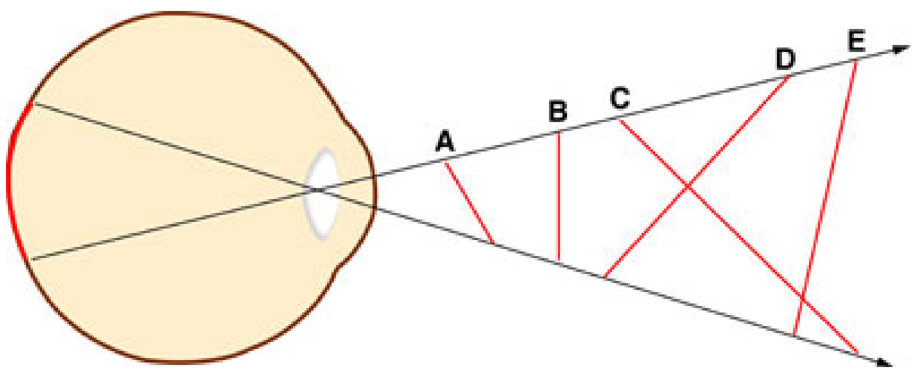


Figure 1.1 The same retinal image may correspond to an infinite number of possible distal stimuli

Yet, we usually have a single and stable conscious perception of the world. One modern approach formulates visual perception as a visual inference problem: a visual system infers a most probable external world through inference. Internal structures and rules in the visual system play important roles in completing missing pieces, restoring depth information, and eventually giving rise to conscious perception. To fully understand this inference process, a first step is to study the representations in the visual system.

CHAPTER ONE

A representation is a system of symbols and rules that describe some content. Different representations render some content explicit and others implicit or hidden. Given one representation of information, some computational problems are easily solvable because algorithms can convert the explicit content into useful solutions. On the other hand, other computational problems may be difficult to solve because useful signals for solutions are embedded in the hidden content of the representation. For example, a binary numeral directly reveals whether a number is a power of 2. In contrast, it requires cumbersome computations to determine whether a number is a power of 10 (Marr, 1982). A frequency representation of a wave signal can easily filter out unwanted frequency components in the signal, but it cannot compute autocorrelations to seek periodic or recurrent behaviors in the signal or to determine temporal synchronizations between this wave and another wave. Many computational problems become easy to tackle when their initial conditions are converted into a new representation. For example, many geometry problems has found simpler solutions after analytical geometry was established and algebra was introduced to transform initial conditions of these geometry problems into algebraic equations. Properties and solutions to large systems of linear equations, which consist of hundreds and thousands of equations, become clearer when matrices are used to represent these systems. All in all, a representation plays a fundamental role in a computational theory. It determines the complexity of the required algorithms that solve computational problems.

This thesis aims to understand position representations in vision. Section 1.1 focuses on defining the scope of our current problem: what problems do we need to address when we study position representations? It discusses two theoretical questions we want to address in this dissertation. First, the definition question asks what set of positions is represented in vision. Second, the format question asks what format the representation uses in its system.

Section 1.2 further discusses the definition question. It aims to identify the represented reference frames in which positions are defined. Two categories of reference frames are considered: an allocentric reference frame characterizes positions in an environment independent of observers; an egocentric reference frame, in contrast, describes positions in relation to the body of observers. Each reference frame has advantages and disadvantages in solving different computational problems and imposes different constraints on the representation format.

Section 1.3 focuses on the format question. It discusses how a representation organizes position information and compares features of different formats. Each format provides easy access to some aspects of position information and supports algorithms to solve computational problems efficiently.

Section 1.1 Two Problems in Position Representation

Marr (1982) reasoned that to understand vision, one level of analysis should focus on identifying computational problems that need to be solved by a visual system. To understand a position representation, we also need to ask the right questions about representation. According to McCloskey (2009), research on a position representation requires understanding how positions are defined as an entity for a representation (the definition question) and what format is used to organize information in the representation (the format question).

Any system that describes positions in space faces the same fundamental definition problem. For example, physics has shown that a single position is well defined only when a reference frame is determined. Many breakthroughs in the history of physics, and of human knowledge in general, have been accompanied by discovering and understanding a counter-intuitive yet important system of reference frames. For instance, the heliocentric view of the universe starts with identifying the solar system as a reference frame centered on the Sun. Newton's mechanic relies on an understanding of the inertial frame of reference. Einstein's relativity theory concerns even more complex reference frames and interactions between them in the four-dimensional space-time model. Daily life descriptions of locations often implicitly assume a frame of reference. Thus, "a person lies on the backseat statically" may have the assumed reference frame of a moving car; this same person may actually be moving at 60 mph if the ground is used as the reference frame.

Implicit assumptions on a reference frame do not usually cause problems in daily communication, but they may prevent experimental investigation from understanding unobservable internal representations in a mind. When a mind controls an agent to interact with environments or other agents, it may represent a position in reference to human bodies in one case, or independent of human bodies in other cases (Klatzky, 1998). For example, the mind can conveniently describe space in a body-centered reference frame to control posture and body movements, because body limbs have limited length and only move in a limited space around the torso. Such an egocentric reference frame, a reference frame that is centered on an agent itself, need not waste time and resources to describe positions whose distances are not reachable. However, position information defined in reference to the body becomes highly inconvenient when the mind needs to guide the body to navigate in the world. Walking in a shopping mall or driving in a large city requires position information far beyond the distance an agent can reach or perceive. In these cases, representing a position that is independent of the agent's body is necessary. Maps are an example of such position representations made by humans. The mind may adaptively use different reference frames in each computational scenario, but this choice is often not consciously accessible by introspection and remains implicit in computations. Therefore, one step to understand a position representation is to explicitly characterize what reference frames are used to define positions in computations.

A system that describes positions also needs to organize information with symbols and rules in the system. Perceptual algorithms that operate on a representation system manipulate its symbols to provide a faithful perception of the external world.

An intuitive format of visual representation is the retinal image. This format describes intensities of each point in image (Marr, 1982), but lacks necessary structures for an accurate inference of the external world. Specifically, an undistorted retinal image does not guarantee a faithful perception of an external position. For example, patients with neurological deficits, such as optic ataxia and William syndrome (Rossetti, Pisella, & Vighetto, 2003; Perenin, & Vighetto, 1988; Landau, & Hoffman, 2012), often fail to use retinal information to localize or relate their bodies to distal objects, even though their visual acuity is normal or corrected-to-normal. Moreover, normal observers also fail to accurately perceive positions in retinal images. They underestimate the turning distance when a dot moves back and forth (Sinico, Parovel, Casco, and Anstis, 2009), perceive positional deviations when a flashing dot is physically aligned with another moving dot (Eagleman & Sejnowski, 2007), and experience spatial compression before they make a saccade (Ross, Morrone, & Burr, 1997). These misperception phenomena suggest that a position representation must have extracted variables from retinal images to describe different aspects of the positions of objects in space.

Formats of different structures extract different sets of variables and use different symbols and rules to organize information. These variables are primitives in a format that explicitly describes a few aspects of a position, and they are easily accessed by algorithms. Algorithms that need to access different aspects of position information must choose different formats to compute efficiently. Thus, in this thesis, we investigate the format of representation by analyzing computational algorithms for visual position inference.

Section 1.2 Reference Frames Used in the Mind

From the perspectives of information theory (Shannon, 2001), a representation system needs to specify a position from a set of possible positions to convey messages of position. Therefore, understanding a position representation system requires identifying the set of positions represented by the system. The set of positions is determined by a reference frame based on a group of invariants that are assumed to be static.

An organism can use two categories of reference frames. First, egocentric reference frames use different parts of the organism as the invariant factor. This invariant factor may define the origin point of a reference frame. For example, a human observer uses an eye-centered reference frame to process visual information (Engel, Glover, & Wandell, 1997), a head-centered reference frame to receive auditory information (Yantis, 2013), and a body-centered reference frame to plan and execute motor actions (Guenther, Bullock, Greve, & Grossberg, 1994). Second, the allocentric reference frames use reference points or objects in the external world as the invariant factors. For example, rodents and humans use multiple landmarks in the environment to define an environment-centered reference frame in navigation (O'Keefe, 1976; O'Keefe & Dostrovsky, 1971). This environment-centered reference frame forms a description of objects and locations in that environment regardless of where an observer stands.

Section 1.2.1 Egocentric Reference Frames

People interact with the world by receiving sensory information and exerting actions. Afferent sensory signals, such as visual, auditory, and tactile signals, converge in the brain via different pathways. Our cognitive system needs to integrate these signals to have a consistent conscious experience, and sends motor signals to muscles to coordinate our body. A set of egocentric reference frames can conveniently support such cross-modality computations. They all use the body to define positions and provide similar structures of position description. Algorithms can quickly read out relative positions to bind cross-modality information.

Vision, for example, mostly uses an eye-centered reference frame to describe position information. This can be observed using afterimage and aftereffects experiments, where the adaption to a visual stimulus leaves a perceptual trace in the visual field that moves with eye movement and interferes with subsequent perception of stimuli in the same retinal position (Mathôt & Theeuwes, 2013; Knapen, Rolfs, Wexler, & Cavanagh, 2009). Similar aftereffects exist when people adapt to a single static dot, motion dots, or faces (Wenderoth, & Wiese, M. 2008; Knapen, Rolfs, Wexler, & Cavanagh, 2009; c.f. Turi & Burr, 2012 Afraz & Cavanagh, 2008, 2009), suggesting that an eye-centered reference frame is used in computations at different levels of visual perception.

Neuroscience also complements evidence for the eye-centered reference hypothesis in vision. Neurophysiological research has shown that spatially adjacent information on the retina activates spatially adjacent neurons in several visual cortices. This correspondence forms a topological mapping relationship, called retinotopic map, between information on the retina and information in the visual cortices (Engel, Glover, & Wandell, 1997). Patient studies (Holmes, 1944), electrophysiological studies (Hubel & Wiesel, 1977), and neuroimaging studies (Engel, Glover, & Wandell, 1997) have shown the retinotopic map structure from the early visual cortices to parietal areas (Bettencourt & Xu, 2016, Schluppeck, Glimcher, & Heeger, 2005; Schluppeck, Curtis, Glimcher, & Heeger, 2006) and frontal areas (Kastner, DeSimone, Konen, Szczepanski, Weiner, & Schneider, 2007). Moreover, retinotopic maps are also found in subcortical structures, such as in the superior colliculus (Katyal, Zughni, Greene, & Ress, 2010; Schneider & Kastner, 2005) and pulvinar (Bender, 1981; Fischer & Whitney, 2012). Similar structures of representation that depends on retinal positions are also found in temporal areas (Fischer, Spotswood, & Whitney, 2011; Golomb & Kanwisher, 2011). How such a retinotopic map is used in many visual computations is not explicitly characterized, but the observation that this correspondence map widely exists in many brain areas does suggest that the visual system processes information in an eye-centered space.

Egocentric reference frames have also been discovered in neural systems that are responsible for sensorimotor control. These systems in general combine information in both eye-centered and head-centered reference frames and map this information onto a body-centered reference frame to plan, coordinate, and execute the movements of different body parts (Guenther, Bullock, Greve, & Grossberg, 1994). For example, when visual information guides arm movements, information in the head-centered reference frame is transformed into a shoulder-centered reference frame (Soechting, Tillery, & Flanders, 1990).

Although egocentric reference frames are useful and widely exist in our cognitive and neural systems, they are not convenient to solve all computational problems. They strongly depend on the observer and mainly represent local information experienced by him or her, and are thus constrained by the scope within which the observer interacts with the environment. Hence, these reference frames lack global, observer-independent descriptions of an environment or object.

Section 1.2.2 Allocentric Reference Frames

Allocentric reference frames are useful to guide organisms around their environment and extrapolate information beyond their sensory experiences. For example, to move around a large city like New York City, an individual has to consult a representation of the city to plan routes. She may find a representation that is independent of her body, such as a map, useful, because she can easily search for routes between any two locations and discover shortcuts without being there. She may also use her own cognitive map to simulate route planning by activating the place cell system in her hippocampus area (Knierim, Kudrimoti, McNaughton, 1996; O'Keefe, 1976; O'Keefe & Dostrovsky, 1971; Tolman, 1948).

When a person needs to recognize an object from multiple viewpoints, under various lighting conditions, or in different shades and scales, she also needs to engage an allocentric representation of the object centered on the object itself. Otherwise, recognizing objects would be impossible considering the complicated visual environments (Marr 1982). Neuropsychological studies have also provided direct evidence of such object-centered representations. McCloskey (2009) reports on a student, AH, who constantly misperceived an object to be its mirror reflection, defined in relation to the object’s main axis. This misperception prevailed in AH’s reading, drawing, pointing, and other tasks (McCloskey, 2009). Furthermore, some visual neglect patients only ignore half of every present object in their whole visual field. They can draw all the objects in a target picture, but they only draw one side of each object, as if the other side does not exist (Driver & Halligan, 1991; Halligan, Fink, Marshall, & Vallar, 2003). These errors in performance confirm that visual recognition must use an object-centered representation that is independent of the observer.

Allocentric reference frames also have disadvantages and may not be suited for several computations. These reference frames need to build up information gradually from multiple pieces of direct sensory information from the egocentric reference frames (Klatzky, 1998). They require complex alignment algorithms and coordinate transformation computations to bind multiple pieces of information that are separated into independent reference frames centered on different objects and environments. Furthermore, computations become more expensive if correspondences need to be dynamically built across time and space while an organism is moving.

Section 1.2.3 One Space in Two Reference Frames

Egocentric and allocentric reference frames coherently interact with each other in the mind and yield a single experience of space. However, they have their own features and problems in describing positions. It is important to understand how they compensate for each other's problems to effectively describe positions in all scenarios, and what constraints they impose on the representational formats and algorithms.

It is difficult to clearly dissociate the roles of the two kinds of reference frame when we try to understand how they complement each other's descriptions of space. Despite different computational complexity, both kinds of reference frames can usually find algorithms to solve the same problems. This leads to debates on whether one kind of reference frame is necessary at all in some cognitive processes. For example, both an allocentric representation and an egocentric representation have been proposed to explain the stable perceptual experience across saccades (Breitmeyer, Kropfl, & Julesz, B, 1982; Cavanagh, Hunt, Afraz, & Rolfs, 2010). Furthermore, emphasis on the computational feasibility of the egocentric representation has called into question the existence of the allocentric representation in the cross-saccade integration process (Cavanagh, Hunt, Afraz, & Rolfs, 2010). Neuropsychological research can be useful in addressing such debates and clearly separating the two kinds of reference frames in cognitive processes. We will examine a case where position information from two reference frames is not consistent in Chapter 2.

Different reference frames also imply constraints on representational format to capture their own features. An egocentric reference frame requires the corresponding representational format to clearly denote the origin point and to structure a position description with an implied relationship to the origin, because it heavily relies on a single invariant factor: the body of an organism. In contrast, an allocentric reference frame requires the corresponding format to not put an emphasis on any point in space and to instead carve the entire space in units of equal size, because it needs to present an undistorted space to the algorithms regardless of where the organism is located or what position is used.

Section 1.3 Formats for Representing Positions

Each position in a reference frame needs to be discriminated from other positions. This concerns a format of a representation that organizes symbols to denote each single position in the set. Different representational formats make certain aspects of position information explicit at the expense of hiding other aspects in the background. The explicit part, encoded in the primitives of the representation, is readily accessible, whereas the hidden part may become difficult to recover (Marr, 1982). Therefore, depending on what primitives have useful information for a computational problem, certain formats can lead to the use of easy and straightforward algorithms to solve the problems, whereas other formats may require complex algorithms, or may find no feasible algorithms, to solve the same problems. For example, as illustrated by Marr (1982), to determine whether a number is a power of 2, a binary format provides a simple and intuitive solution: any number with a single 1 and 0 or multiple 0s, such as 1000 (in binary format), is a power of 2. Solutions in a decimal format are more difficult: we need to recursively test whether a number is 1 or whether half of its value is a factor of 2. Moreover, a solution based on the Roman or Chinese numeral is difficult to conceive. If it exists at all, the solution cannot be as simple as the previous ones. Similar logic applies to computations in visual systems. What format is used to describe a position can affect how the visual system solves computational problems that involve position information.

Three candidate formats have been suggested in a cognitive system: a propositional one, a nominal one, and the coordinate system. They appear in different species and different developmental stages. To understand which candidate is used in vision, we need to study both the characteristics of each format and their constraints and requirements in visual computations.

Section 1.3.1 The Propositional Representation

We use languages to describe positions. In daily conversations, we do this by saying that *this object is ON TOP OF the table*. Similarly, we describe that *a chair is NEXT TO a desk*, *the desk is TO THE LEFT OF a lamp*, and *the lamp is BELOW the ceiling of the room*. That is, we use a sequence of propositions to describe locations in space. This sequence of propositions is constructed by organizing variables, which denote objects (e.g. WHITE WALL), and predicates, which indicate spatial relationships (e.g. ON TOP OF), in a specific syntactic structure. By using these propositions, we can describe and understand spatial relationships between multiple objects and specify a position by performing the predicate calculus. One advantage of the propositional representation is that it offers a compact and robust description of spatial relationships even if large position perturbations are introduced. For example, an apple that is ON TOP OF a table could change its location in space dramatically, such as from the left corner to the right corner of the table surface, but still remain ON TOP OF the table. This robust propositional representation is useful to maintain an understanding of multiple positions and their relationships in the presence of perceptual noise.

Developmental research has provided insightful evidence suggesting that people indeed use a propositional representation of location to solve computational problems, such as searching in space after disorientation (Hermer & Spelke, 1996; Hermer-Vazquez, Spelke & Katsnelson, 1999) or binding features into objects (Dessalegn & Landau, 2008; Landau & Lakusta, 2009). For example, four-year-old children usually mismatch a bi-color target with its mirror reflection counterpart in a delayed-to-match task. A spatial proposition, such as "the red is to the left of the green", can improve children's matching accuracy from 60% to 80%, but the similar instruction "the red is touching the green" does not help children in this task. Therefore, the directional label [*TO THE LEFT OF*] in the proposition is the key that helps children to process position information. Similar results from both adults and children suggest that humans use a propositional format to represent a location; in turn, this propositional representation helps other non-linguistic representations in solving visual tasks (Landau & Lakusta, 2009).

Section 1.3.2 The Nominal Representation

A nominal format uses labels, or names, to denote different locations (McCloskey, 2009). It can clearly discriminate between different locations and provide a concise guide for the navigation or path integration. When a new location needs a denotation, we create a new symbol to name this location. For example, we use New York City to represent the area of a city located at the southern tip of the state of New York. By labeling different locations, we can easily determine whether the current location is the same as a previous one. We can also make an ordered list of labels, such as *(Penn, Aberdeen, Bowie State, BWI airport)* to construct a path or a trajectory of motion, as in a subway map.

Neuroscience research on the hippocampus system has found a place cell system that represents positions in a way similar to a nominal format. If a rat is freely moving in a bounded area, some pyramidal neurons in the dorsal portion of its hippocampus, CA1, will become active when the rat enters a particular place in that area. This place in the environment is called the place field of the activated neuron, and the neuron, because it is sensitive to a specific place, is called a place cell (O'Keefe and Dostrovsky, 1971). A neuron is activated, and the rat's brain represents the position where the rat is located in the environment at that moment. The activities of a population of place cells thus create a neural map representing the current external environment (O'Keefe, 1976; O'Keefe and Conway, 1978). Such a neural map is suggested to be a neural correlate of the cognitive map (Tolman, 1948), based on which rats navigate in environments. Importantly, the place cell system does not resemble the environment as a map does. Neighboring place cells may represent nearby fields or distant ones in equal probability (O'Keefe, Burgess, Donnett, Jeffery, & Maguire, 1998). Thus, neurons indeed act as independent labels. They also only reflect the order of positions on a path and ignore the geometry in the environment (Dabaghian, Brandt, & Frank, 2014). Thus, a sequence of neuron firings represents a path, similar to how a subway map does.

Section 1.3.3 The Coordinate Representation

A coordinate system uses an ordered tuple of numbers, called coordinates, to describe a position in a reference frame. It is a compositional and quantitative representational format that can easily describe positions in any multi-dimensional space. For example, a two-dimensional coordinate system may describe a position in the form of (x, y), which specifies this position's signed horizontal and vertical displacement from two cardinal axes in the reference frame. Coordinate systems link geometry to algebra and provide analytical solutions to geometry problems.

Cartesian Coordinate System

The Cartesian coordinate system has straight, perpendicular axes that partition the space into rectangular grids. Each coordinate represents a signed distance to one axis. Each rectangular grid has the same area regardless of where the origin point is located. As a result, the Cartesian coordinate system does not heavily rely on the origin point and carves the space equally everywhere. Furthermore, translations and reflections are simple in the Cartesian coordinate system. The former require only a linear calculation, and the latter a simple sign change in coordinates. The Cartesian coordinate system is a natural and intuitive choice when researchers study position representations.

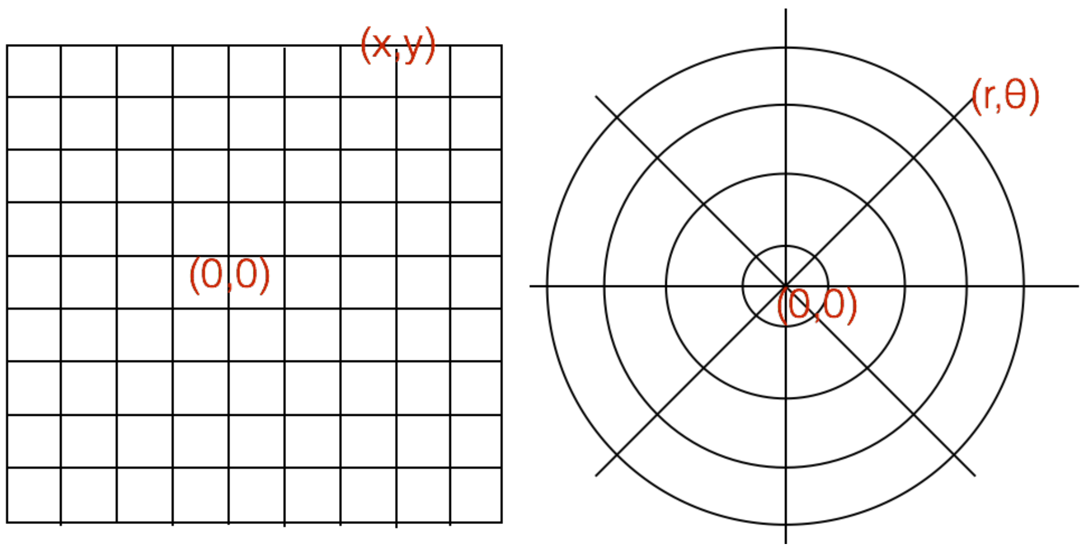


Figure 1.2 The Cartesian coordinate system and the polar coordinate system

Artificial visual systems, such as computer vision systems, mainly use the Cartesian coordinate system. This is a reasonable engineering choice because most inputs are in rectangle shapes, such as images, photos, or papers. These inputs usually have information organized in row-by-column matrices. Computations, such as edge detection, blurring and sharpening, and optic-flow motion analysis, are usually calculated with respect to row information (x-axis) or column information (y-axis).

When it comes to the human cognitive system, Gallistel (1990) argued that the Cartesian coordinate system can avoid accumulating errors in successive small movements and can prevent carrying errors to future position computations because it partitions the whole space into equal-sized rectangular grids. This is useful in navigation (Gallistel & Matzel, 2013; Klatzky, 1998), and could also be useful in other cognitive processes that involve small, successive changes in space. For example, in the visual domain, the saccade system seems to operate in a Cartesian coordinate system. Moreover, physiological and behavioral evidence has shown that the saccade system has independent horizontal and vertical components (Bahill and Stark, 1977). Electrophysiological studies have identified separate neural channels with different motor control centers, different motor neurons, and different eye muscles that implement the horizontal and vertical components of a saccade respectively (Bahill and Stark, 1975, 1979). Behavioral studies have also shown that asynchrony in vertical and horizontal subcomponents of oblique saccades leads to curvature in saccade trajectories (Bahill & Stark, 1975).

The Curvilinear Coordinate System

The curvilinear coordinate system includes a polar coordinate system of two-dimensional planes, and spherical and cylindrical coordinate systems of three-dimensional space. All of these specify a position in space in terms of distances and angular bearings to the origin point. As a result, a curvilinear coordinate system can easily represent rotations and curves, such as circles and spirals. Unlike the Cartesian system, this system places more emphasis on the origin point and partitions the entire space into unequal unit grids – the farther away the space is to the origin, the larger the unit grid becomes (Figure 1.2).

Abundant neuroscience evidence suggests that the brain system uses curvilinear coordinate representations to process sensory and motor information. Early visual cortices have been suggested to use a logarithmic-polar coordinate transformation of retinal images. Holmes (1944), who established the organization of the V1 by examining injured veterans, first found that information in the occipital lobe seems to require a polar coordinate transformation of information on the retina: information along the ventral-dorsal axis corresponds to information at the same distance to the fovea from the upper visual field to the lower visual field; while information along the posterior-anterior axis represents information at the same angle from the fixation from the fovea to the periphery. In addition, cortical representation of the fovea occupies a larger area in V1 than cortical representation of the peripheral visual fields does, suggesting that the cortical representation partitions the visual fields with unequal units (Coway and Rolls, 1974).

Schwartz (1980a) established the mathematical formulation of the logarithmic-polar coordinate transformation computation in V1. This log-polar neural mapping of information in the visual field is found in primary visual cortices in different species, such as cats, monkeys, and humans, with differences only in the values of parameters (Schwartz, 1980a, 1980b, 1983). Mathematical modeling of geometrical visual hallucination has also shown that a log-polar coordinate system in V1 provides the best explanation for such visual phenomena (Ermentrout & Cowan, 1979; Bressloff, Cowan, Golubitsky, Thomas, Wiener, 2001).

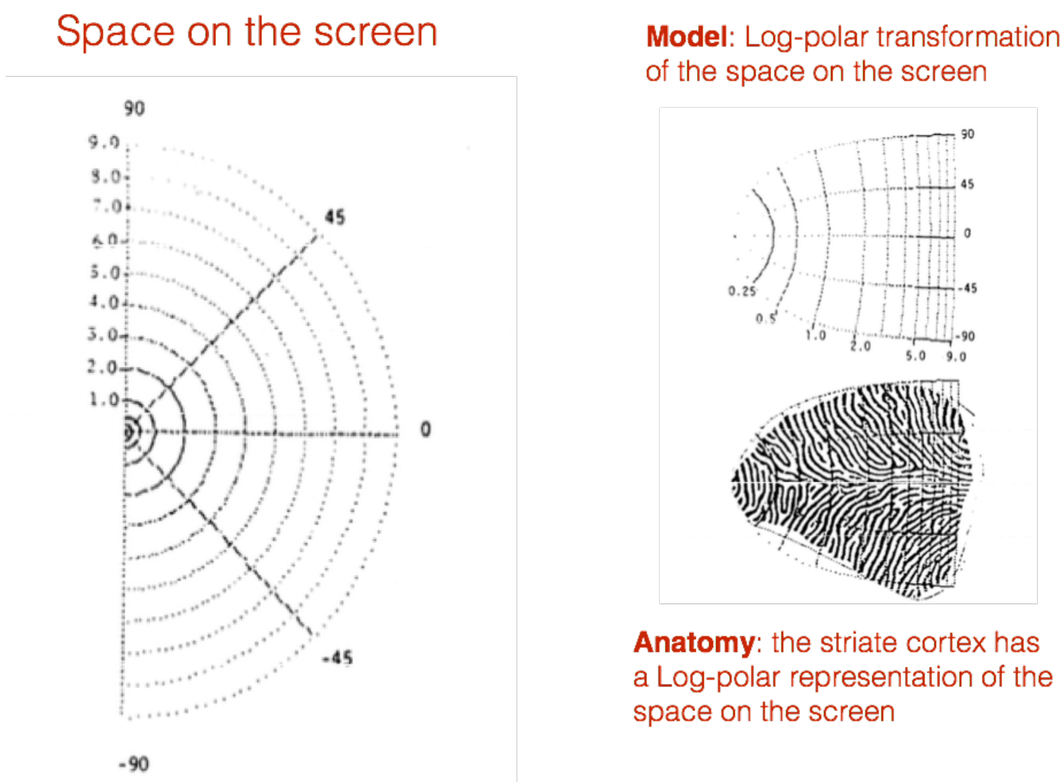


Figure 1.3 Log-polar coordinate system in the primary visual cortex (Schwartz, 1983)

Although more controversial, evidence also suggest that a polar topological mapping of information in visual fields extends beyond V1. In the dorsal pathway, retinotopic maps as far as in the parietal cortices can be segmented using angular rotation (Silver & Kastner, 2009). In the ventral pathway, Levy and colleagues (2001) have shown that different ventral cortices have different biases in processing information from the fovea and the visual periphery, suggesting an unequal partition of visual space in the ventral pathway.

Curvilinear coordinate systems are also found in other sensorimotor areas. Neurons in premotor areas, motor areas, and visuomotor areas in parietal cortices encode preferred directions and distances in planning and executing motor signals (Caminiti, Johnson, Galli, Ferraina, & Y. Burnod, 1991, Lacquaniti, Guigon, Bianchi, Ferraina, & Caminiti, 1995). Even in subcortical areas, such as the superior colliculus (SC), curvilinear coordinate systems are found in multiple layers to align visual and auditory signals and to compute a master salient map that plans motor responses, such as saccades (Knudsen, 1982; Knudsen & Brainard, 1995; Wurtz & Goldberg, 1972).

This prevalence of curvilinear coordinate systems in sensory and motor brain areas may be closely related to the choice of an egocentric reference frame. As discussed in Section 1.2.3, egocentric reference frames require a clearly defined origin. From the origin, a body can only stretch out and rotate within a limited distance. A three-dimensional spherical coordinate representation can readily characterize the space accessible to the body in the egocentric reference frame. Different sensory and motor signals can also accurately coordinate information by using the curvilinear coordinate systems that either describe a two-dimensional space (like the visual field) or a three-dimensional space (like the motor space). The computational difficulties in the mapping process can be best exemplified by the geological information system (GIS) design: when representing the surface of the Earth, a two-dimensional map rarely uses a Cartesian coordinate system. Instead, the GIS uses a longitude-latitude system or a polar coordinate system to present the surface of the Earth from different viewpoints.

Section 1.3.4 A Probable Representational Format

Representations serve to feed input to algorithms (Marr, 1982). The representational format that is likely used in vision should be reflected by algorithms in visual perception. Visual perception requires a wide range of quantitative computations to calculate motion, speed, distance, length, etc. The propositional and nominal representations do not have the primitives that easily encode quantitative information about positions. In contrast, the coordinate system has multiple computational advantages.

First, a coordinate system provides metrics with which one can measure magnitudes such as length and size. Each component of a coordinate system has a zero value and a unit magnitude, thus it can afford algebraic computations such as addition, subtraction, multiplication, and division. These basic algebraic computations can support complex perceptual computations, such as calculating the distance between two locations, translating or rotating a point from one location to another, or scaling or shearing aspects of a shape in the space. Considering all complicated geometric transformations in the visual system that involve positions, only a coordinate system can make all required quantities available for each calculation.

Second, with its compositional nature, a coordinate representation efficiently encodes position information with a few primitives. Each primitive encodes one aspect of the position independent of other primitives. Thus, a coordinate representation requires much fewer symbols to represent a multi-dimensional location than other formats. For example, to represent locations in an N-by-N grid system on a two-dimensional plane, a nominal representation must use N^2 symbols to distinctively represent each location in this grid system, but a two-dimensional coordinate system only needs 2N symbols. When number of dimensionality or the scale of each dimension increases, the number of symbols in a nominal representation may increase exponentially, whereas the number of symbols in a coordinate system only increases linearly. Thus, the coordinate system can represent a much larger scale of space without suffering from the curse of dimensionality.

Therefore, in our empirical studies, we focused on understanding coordinate systems in vision.

Section 1.4 Organizations of Empirical Studies

As part of this thesis, we have conducted empirical studies to examine both the definition question and the format question regarding position representations in vision. In Chapter 2, we present our exploratory neuropsychological case study of a posterior atrophy patient. We used a series of testing to identify different reference frames used by the patient to define positions in his visual perception and visually guided actions. We also tried to understand the format of position representations in his eye movement system by studying his saccade trajectories. These attempts inspired three subsequent studies that focused on examining the format of position representation in various visual domains, including perception, visually guided action, and eye movements.

To examine the format of position representations, we adopted a hypothesis-driven approach. We asked: when and what coordinate system is used in position representation? We tested models of two alternative hypothetical formats: a Cartesian coordinate system and a polar coordinate system. In our empirical studies, we always started by deriving expected behaviors based on these two models. We then designed behavior experiments, collected data from participants, and inferred which model was more likely to be used in vision. Next, in Chapter 3, we reason that Cartesian and polar models predict different shapes of noise distributions, so we model distributions of various behaviors and identify the model that can better explain the variance and bias in the behavior responses. In Chapter 4, we argue that primitives in Cartesian and polar models are different. A Cartesian model has primitives as horizontal and vertical translation magnitudes, whereas a polar model has primitives as distances and angles to a fixed origin. We examined which primitives are used in vision by asking participants to compare angles and distances, and to align dots on straight lines. In Chapter 5, we then posit that computational complexity is affected by representation formats and show that the same geometrical feature can either have a simple linear description or a complex non-linear description, depending on what format is used to describe positions. We asked participants to recognize patterns among noise, and measured how difficult it was for them to perceive different global geometrical patterns. The perceptual difficulty should reflect the computational complexity of perception algorithms, and in turn reflect what coordinate system is used to represent positions in the first place.

Chapter 2 MDK Case Study: Misperceiving Positions in Space

Neuropsychology has provided many opportunities for us to understand the human brain and cognition. Neurological deviation from normal experience usually leads to surprising observations of human behavior, which in turn drive experimental research to investigate underlying neural or cognitive mechanisms. The patient HM, for example, led to an era of research in learning and memory (Corkin, 1984). Similarly, Patient Gage revealed a connection between the frontal lobe and personality (Gazzaniga, 2004), and Patient DF helped us to understand a dissociation between two information pathways in the visual system (Goodale, Milner, Jakobson, & Carey1991). Moreover, agnosia patients drove research on object representation and the functions of visual cortices (Benson, & Greenberg, 1969). These opportunistic case studies often reveal unanticipated phenomena, challenge existing theories, and yield new theoretical insights (McCloskey, 2009).

Since 2015, we have had the opportunity to test a posterior cortical atrophy patient, MDK. MDK initially reported a reading disorder and a blind spot in his visual field. However, our testing soon revealed that his major deficits are in his spatial representations: he fails to point to a peripheral location and his errors are large and systematic. Here, we present our investigations of MDK to shed light on questions of position representation, including the reference frame and the representational format.

Testing MDK is an exploration; hence, exploratory results are foreshadowed here. We first examined what reference frame is used in MDK's visually guided actions. Using single dots to denote positions, we showed that MDK's pointing responses were strongly biased towards his fixation point, suggesting that his visually guided actions use an eye-centered reference frame. Accordingly, when his fixation shifted, MDK's pointing responses shifted in a similar way. We then explored how positions of parts of an object are represented. These positions may be independent, similar to a set of dots; or alternatively, there may be an integrated representation of the object, and those positions may be described in a reference frame defined within the object. To examine this, we asked MDK to point to two positions in the periphery. These positions were cued by two separate dots or by a connected bar. We found that MDK's bimanual bar pointing results could not be explained by simply combining unimanual responses from each hand. For instance, his bimanual pointing distance was constantly larger than the sum of his unimanual pointing distances. This non-linear integration from unimanual to bimanual pointing suggests that MDK involves an object-centered reference frame to calculate positions when he points to endpoints of a bar.

In addition, we studied how MDK's eye-centered reference frame is updated when he moves his eyes. Importantly, we investigated how his fixation moves from an old position to a new one, and whether its movement signals are represented by a polar coordinate system or a Cartesian coordinate system. To answer this representational format question, we studied MDK's saccade trajectories. When he saccaded to peripheral target locations, his saccade trajectories implied that his eye movement signals were scheduled in a Cartesian coordinate system instead of a polar coordinate system.

Case Description

MDK is a 68-year-old right-handed male with a PhD degree. He is diagnosed with posterior cortical atrophy and has shown symptoms since 2010, with no notable pre-deficit medical history. His progressive neural degeneration starts in his occipital lobe. A 2012 MRI result shows mild bilateral atrophy in his posterior parietal and occipital cortex, and recent anatomical MRI images (2017) show similar bilateral atrophy, but no more significant deterioration.

2.1 The Reference Frame of Position Representation

In this section, we explore which reference frames MDK uses in his visually guided actions.

2.1.1 The Center of Egocentric Reference Frames

After testing MDK during several sessions, the following was a consistent observation: when he tried to point to a peripheral position while maintaining his fixation, his responses were shifted towards a single origin point: his fixation. Our first question is: what is the center, or the origin point, of the reference frame that defines his response positions?

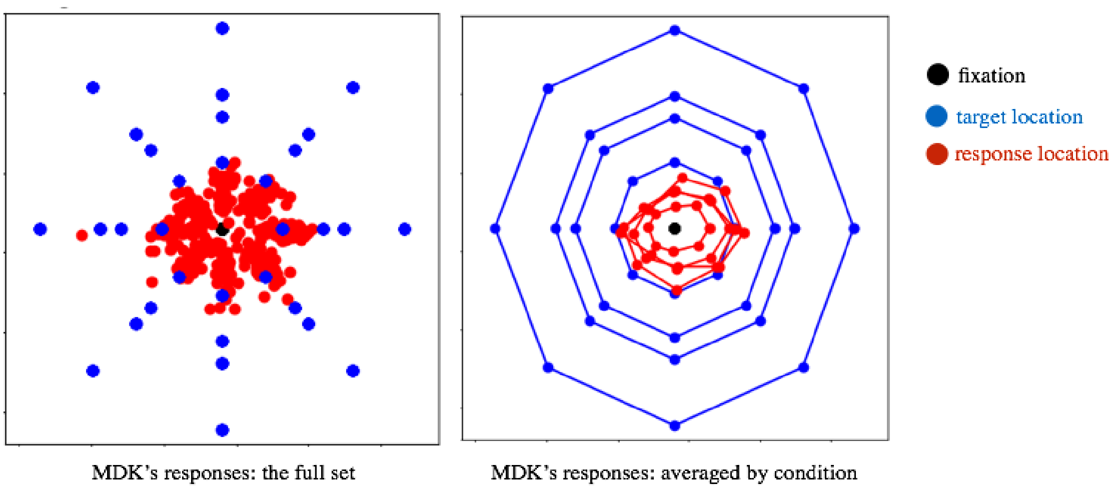


Figure 2-1 MDK's responses to peripheral positions

Usually, MDK directly faced a screen during testing sessions. The center of the screen was the same as his fixation and the center of his head; hence, the center of the reference frame could be any of these three possibilities. We designed tasks to differentiate between these three by asking MDK to fixate on different locations on the screen while presenting stimuli at the same set of target positions across different fixation conditions.

Methods

Apparatus

Stimuli were presented on a Macintosh iMAC computer with a refresh rate of 60Hz. The viewing distance was approximately 40 cm so that the display subtended 66.45° x 41.34° of visual angle. A chinrest was used to stable MDK's head and to ensure that MDK directly faced the screen.

Stimuli

We used a single dot to denote a position in each trial. This dot extended 0.5°. Target positions were at 9 ° or 12° away from the screen center, appearing at one of the eight directions from 0 ° to 315 ° with a step size of 45°.

Procedure

Each experiment trial started with a fixation mark on the screen. MDK was instructed to maintain fixation on the mark and to be ready for visual stimuli. Once MDK indicated that he was ready, a dot flashed at one location on the screen for 200 ms. MDK was required to respond to the location by pointing to it on the screen. MDK was asked to maintain his fixation during his pointing.

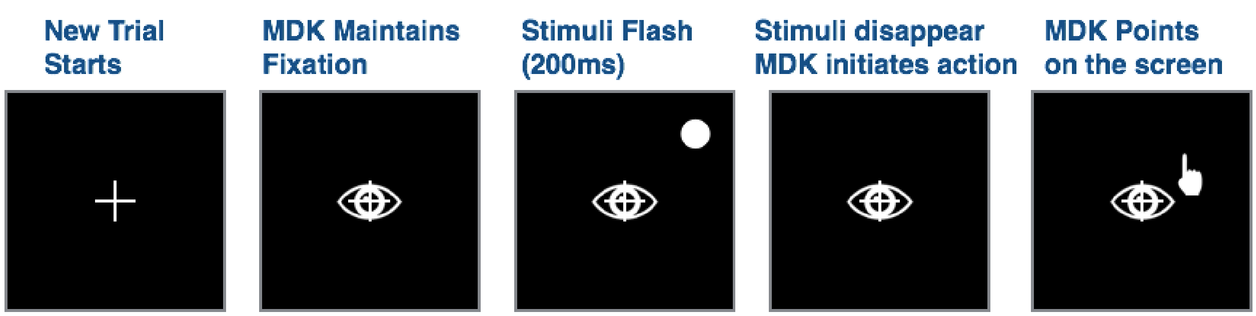


Figure 2-2 General procedure of MDK's pointing experiment

The fixation mark was placed at three different positions on the screen: one appeared at the center of the screen, and the other two appeared on the left or right of the screen center, respectively, with a distance of 10.5°. We further included a closed-eye condition. In this condition, after stimuli presentation, MDK was instructed to point to stimuli with his eyes closed. We asked him not to move his eyes when he closed them.

Results and Discussion

MDK's responses to the same set of target positions on the screen differed as a function of fixation (Figure 2.3). These results clearly showed that his pointing was biased by where he looked. When he looked at the left side of the screen, all of his pointing destinations were biased towards the left side of the screen; and when he looked at the right side, his pointing destinations shifted to the right. In contrast, when he was asked to point to the fixation mark, he accurately pointed to the mark regardless of where the mark appeared. Altogether, MDK's performance shows that his visually guided action represents positions in an eye-centered reference frame.

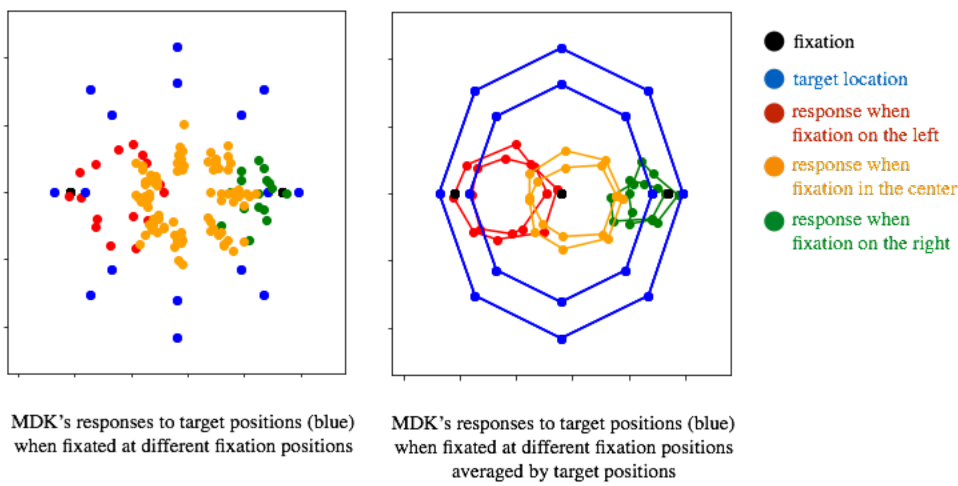
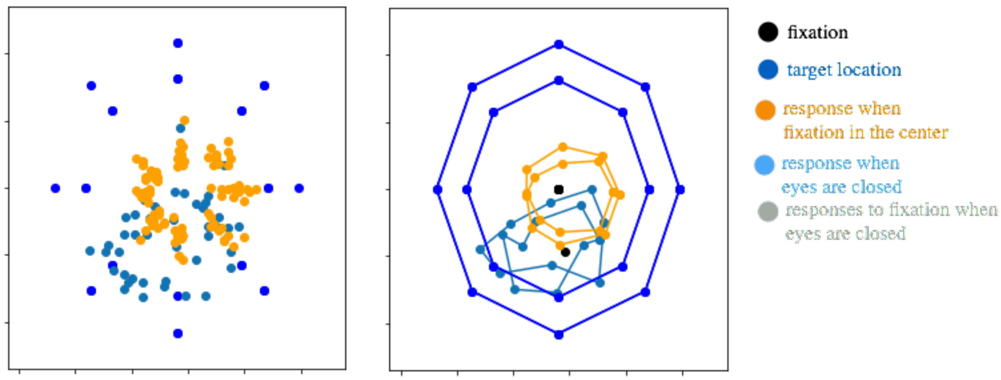


Figure 2-3 MDK's pointing responses were biased towards where he fixated

Next, we studied how MDK maintained such an eye-centered reference frame. In the closed-eye condition, when MDK pointed to peripheral stimuli with his eyes closed, his pointing shifted down overall (Figure 2.4). Even when he pointed to the fixation mark, he consistently pointed to a point lower than the true position on the screen. His pointing positions also did not maintain angle bearing as accurately as they did in previous conditions, even after we corrected his closed-eye pointing responses to the new fixation mark.



MDK's responses to target positions (blue) when his eyes are open or closed

MDK's responses to target positions (blue) when his eyes are open or closed averaged by target positions

Figure 2-4 MDK's pointing responses when his eyes are open or closed

Taken together, our results suggest that MDK's visually guided actions used an eye-centered reference frame to describe positions in his visual fields. This eye-centered reference frame may be maintained by continuous feedback loop between vision and action. However, when this feedback was disrupted by MDK closing his eyes, his closed-eye pointing responses lost track of the positions in the eye-centered reference frame that should have been aligned with the center of the screen. As a result, he consistently pointed to positions that were below the true positions. Downshift biases in MDK's closed-eye pointing may reflect that MDK planned action landing positions in a new egocentric reference frame. Previous electrophysiological research found that position information was transformed between a head-centered reference frame and a shoulder-centered reference frame in visually guided arm movements (Guenther, Bullock, Greve, & Grossberg, 1994; Soechting, Tillery, & Flanders, 1990). When position information was shifted from the eye-centered reference frame to the shoulder-centered reference and no visual feedback was available to re-define positions in the eye-centered reference frame, MDK might have executed his pointing responses in a shoulder-centered reference frame. The center of that frame was below the center of the eye-centered reference frame. Thus, MDK showed an overall downshift bias in his closed-eye pointing.

2.1.2 Egocentric Reference Frames versus Allocentric Reference Frames

Above, we saw that when MDK pointed to peripheral positions, he pointed to a destination that shifted towards a single origin point. The position seemed to approach the point where he fixated. It is possible that his spatial distortion combined scattered objects in the periphery so that those objects crowded and congested around the fixation. This hypothesis is in line with his description of his reading experiences: he reported that he felt words were overlapping, cluttering his visual field.

This hypothesis raises the question of how he can interact with an object, especially one that takes up a large space, rather than a single dot. If the way he perceives an object is the same as the way he perceives single positions, we would expect that he should consistently underestimate the size and length of objects, and thus have difficulties in grasping or reaching for objects properly. However, in our informal interactions with MDK, we found that he could estimate the length of pens and his estimations, when presented to us with his two hands, were not the underestimations that we would expect from his pointing results. Therefore, we then tested him on pointing to oriented bars with a single hand or both hands while monitoring his fixation.

General Method

Apparatus

Stimuli were presented on a monitor that was viewed by MDK at a 40 cm distance. The screen extended 68.04° x 37.80° and refreshed at 60 Hz. A chinrest was used to stabilize MDK's head and to ensure that MDK directly faced the screen. We also monitored MDK's gaze with an SR

Research Eyelink 1000 Tower Mount eye tracker at a 500 Hz sampling rate. A five-point calibration and a validation procedure were run before each testing block. This procedure was repeated until we could accurately track MDK's eyes with a mean error of less than 0.5°. The eye tracker was controlled by in-house MATLAB scripts.

Stimuli

Two kinds of stimuli were used in our testing. One kind was a single dot, which was always 1°. The other stimuli were rectangular bars. The bars had various lengths and orientations on the screen, but they all had the same width of 0.67°.

General Procedure

The experimental procedures were similar to those described in Section 2.1. Each trial started with a fixation mark. MDK fixated on the mark and indicated when he was ready. Depending on the experimental condition, one dot, two dots, or a rectangular bar flashed on the screen for 200 ms. MDK pointed to dots or endpoints of the bar with one hand or two hands.

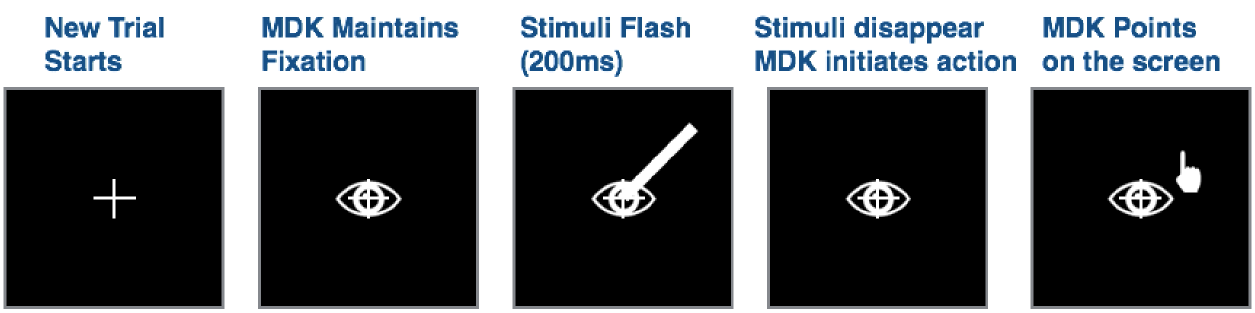


Figure 2-5 General procedure of MDK's bar-end pointing experiment

Experiment 2.1.2a A Single Dot versus a Single Endpoint of a Bar

Is MDK's pointing to the endpoint of a bar the same as his pointing to a dot in the same position?

Method

Stimuli, Design, and Procedure

In this experiment, a single dot or bar flashed on the screen. The bar was always extended from the fixation mark, so it had one endpoint attached to the mark and the other going out in the peripheral visual field. Possible positions of outgoing endpoints were matched with possible positions of the dots. In a reference frame defined within the computer screen, these two sets of positions were the same. MDK was instructed to point to the position of the dot or the outgoing endpoint of the bar. He was instructed to use either his left hand or his right hand to point, with a counterbalanced order across testing blocks.

Results and Discussion

MDK's pointing distance was closer to his fixation when the position was denoted by a dot than when the position was denoted by an endpoint of a bar. When the target positions were at different distances from the fixation, MDK's pointing distance also varied. Figure 2.6 shows MDK's pointing distance at different target distance positions. We fitted a simple linear model to describe the relationship between MDK's pointing distance and the true target distance. Model fitting results showed that MDK's pointing distance was closer to the true distance, thus less biased by his fixation, when he pointed to an endpoint of a bar than when he pointed to a peripheral dot. Furthermore, the results showed that MDK's left hand performance was better than his right hand performance.

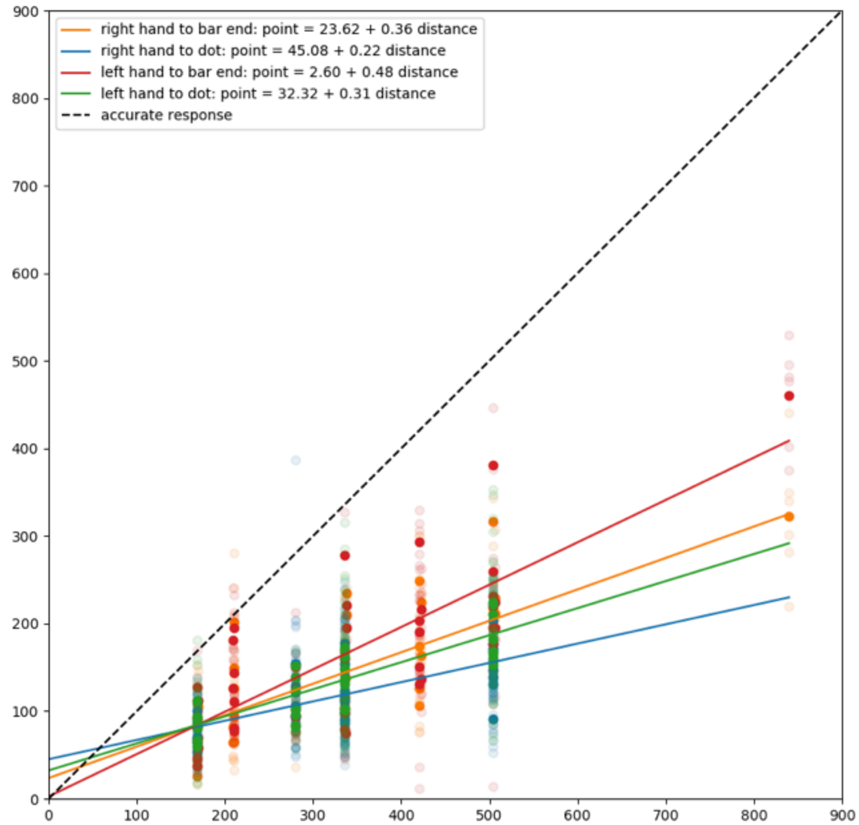


Figure 2-6 MDK's pointing distance when he pointed to a bar or a dot with one hand

Differences between MDK's pointing responses to a dot and to a bar suggest that MDK engaged different information to plan and execute his actions in these two conditions. His better performance in the bar pointing condition suggests that MDK used more information during the bar pointing that was not available in the dot pointing: namely, length.

Experiment 2.1.2b Unimanual versus Bimanual Pointing to Bar Endpoints

We found that MDK's pointing to the endpoint of a bar was better than his pointing to a single dot at the same location, suggesting that he may use length information about the bar to point to the bar. Here, we further tested the possibility that an independent representation of the bar may be used when MDK executed his pointing actions in addition to simple positions in space. In this experiment, we presented MDK with bars of different lengths and orientations and asked him to either point to both endpoints of a bar with two hands, or to point to one endpoint of a bar with one hand. If no independent object representation was involved in his action execution, we would expect two-handed pointing to map onto the addition of pointing results from his left hand and from his right hand. Alternatively, if an object-centered representation is involved, we would expect his two-handed pointing performance to be different from a simple combination of his one-handed performance: information from the object-centered representation would also be integrated in his two-handed pointing and help him guide his pointing.

Method

Stimuli, Design, and Procedure

In this experiment, a bar flashed on the screen. The bar was symmetrical across the fixation mark. MDK was instructed to point to one endpoint of the bar or to both endpoints of the bar. When he pointed to one endpoint, he was instructed to use either his left hand or his right hand, counterbalanced across testing blocks. When he pointed to both endpoints, he was instructed to use both hands simultaneously.

Results and Discussion

MDK pointed to the endpoints of a bar either bimanually or unimanually in the experiment. Our results, illustrated in Figure 2.7, clearly show that MDK's bimanual responses were not a simple combination of two unimanual responses. First, MDK's unimanual responses were closer to his fixation than his bimanual responses were; the response space in the unimanual condition was more contracted towards the fixation than the response space in the bimanual condition. Second, some qualitative differences were observed between his unimanual and bimanual responses.

Closely observe the left part of MDK's bimanual responses (Figure 2.7). His left side responses often overestimated the distance of the true position. In several conditions, his left hand pointed farther than the left endpoint of the bar in bimanual responses. This pattern was in contrast with his highly contracted, fixation-biased unimanual responses. Since the visual presentations were the same in both bimanual and unimanual conditions, MDK's response differences suggest that he further engaged object information, such as the length of the bar, in planning his pointing behaviors.

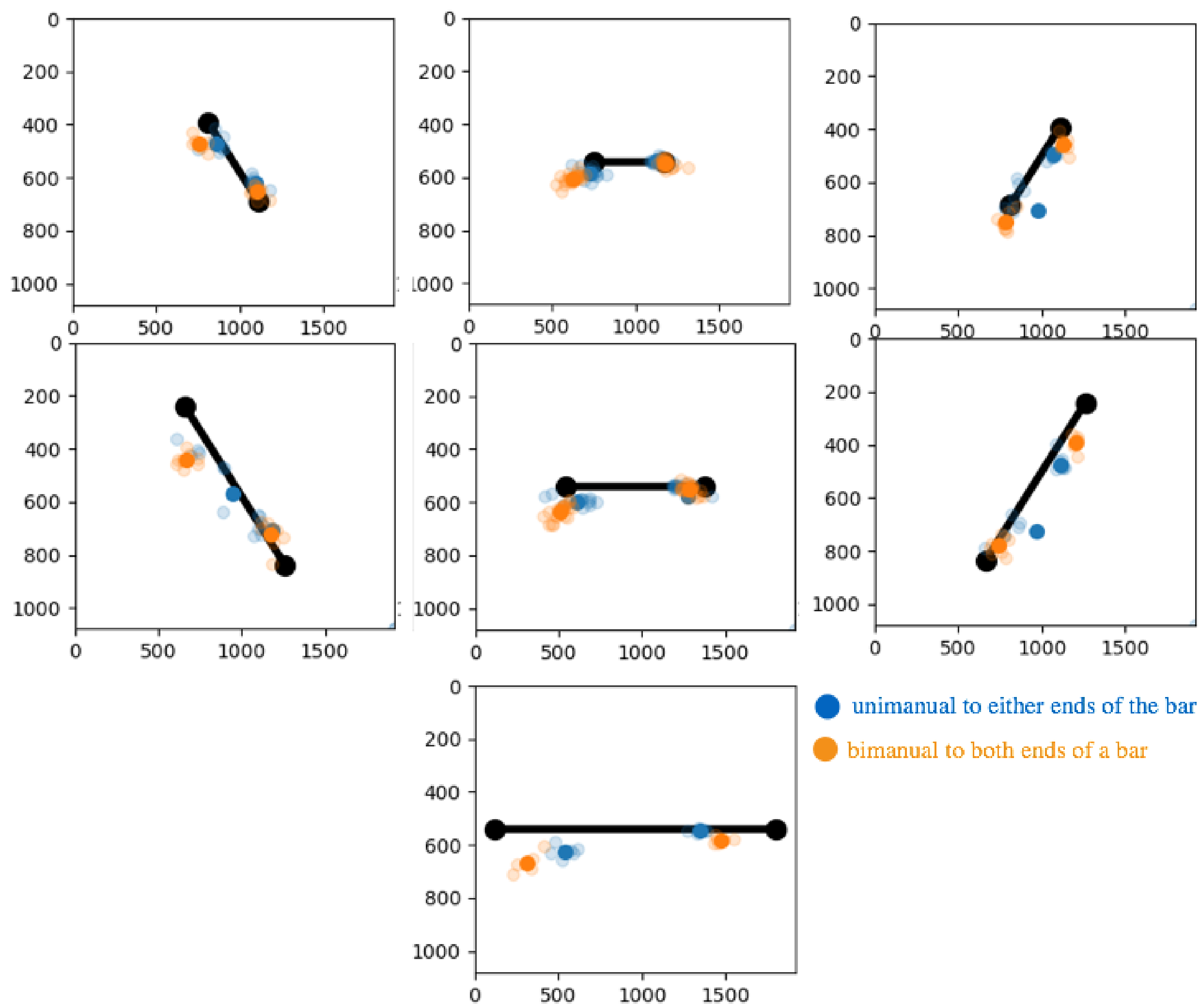


Figure 2-7 MDK's bimanual pointing responses versus unimanual pointing responses

Experiment 2.1.2c Bimanual pointing to bar endpoints and two dots

We found that when MDK used both hands, his responses could not be predicted by the addition of each hand. In this third section, we aim to confirm our hypothesis of an independent object representation. A competing hypothesis may be that differences between two-hand and one-hand pointing were simply due to bimanual coordination in actions. When two hands were in action, MDK could correct one hand's position by referring to the position of the other hand, showing a different pointing response. Therefore, we asked MDK to bimanually point to two dots versus two endpoints of a bar.

Method

Stimuli, Design, and Procedure

In this experiment, two dots or a bar flashed on the screen. The dots were centrally symmetric about the fixation mark, and the same was true for the bar. Possible positions of outgoing endpoints were matched with possible positions of dots. MDK was instructed to point to both positions simultaneously with two hands.

Results and Discussion

In this experiment, MDK pointed to two dots or two endpoints of a bar with both hands. We matched the positions of the two dots and the endpoints of the bar; the only difference between the two conditions was whether two target positions were independent of each other (in the two-dot condition) or connected with an intact object (in the bar condition). If MDK's bimanual performance in the previous experiment was simply due to a bimanual coordination, we would expect him to point to the same positions in this experiment. Alternatively, if he engaged extra information about the bar, we would expect him to point to different positions in this experiment. Figure 2.8 shows MDK's pointing results. When he pointed to the two dots, his responses were simultaneously biased towards his fixation. When he pointed to the two endpoints of a bar, in contrast, his responses were less biased towards his fixation. In several cases, he overestimated the true positions again: his left hand pointed farther than the true endpoint on the screen. His right-hand performances were similar in the two conditions: sometimes the distribution of two groups of responses overlapped entirely. Together, our results suggest that MDK used the visible length of the bar to plan his pointing behavior. It is likely that he used extra information to adjust his left hand to compensate for his persistent fixation biases in his right hand so that his two-hand distance would be better matched to the length of the bar. Such adjustment seemed to be unconscious: MDK never reported any intention of adjusting his left hand. In addition, this adjustment required intact object information, since pointing to two dots simultaneously did not induce similar compensation. Our results further suggest that MDK used an object-centered representation to calculate position information and to guide pointing.

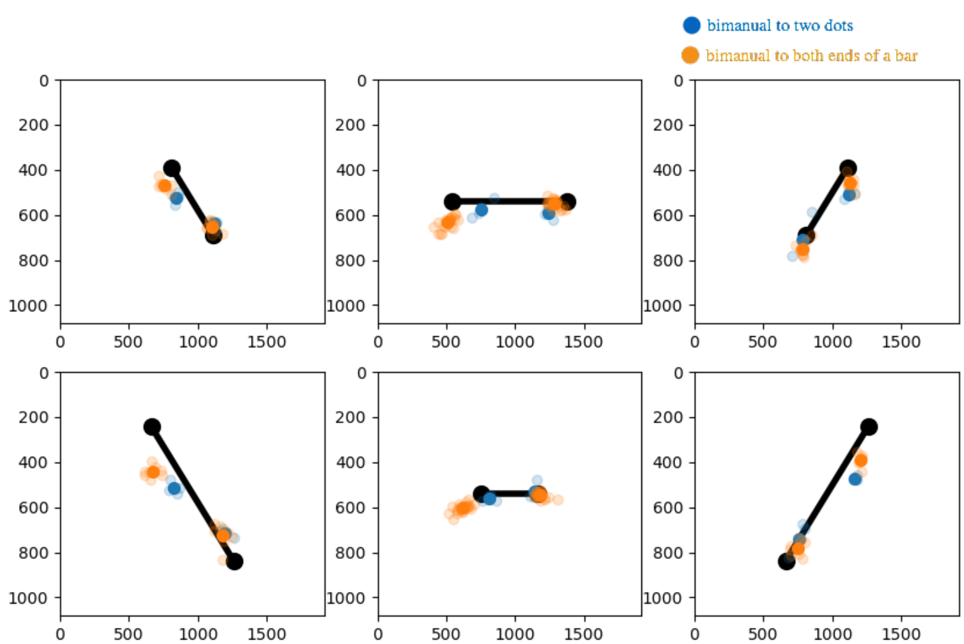


Figure 2-8 MDK’s bimanual pointing responses to two dots versus two endpoints of a bar

2.2. The Format of Position Representation

MDK's pointing responses were systematically biased towards the center, namely, the fixation point, of an eye-centered reference frame. In those cases, the center of the reference frame was static and the reference frame itself did not move.

What format did MDK's position representation use to describe each position in the reference frame? His responses usually preserved the angle information of peripheral dots but drastically failed to reproduce the distance information. Such results seem to suggest that the reference frame has a polar coordinate system where each position in defined as a distance and an angular bearing away from the fixation. MDK's pointing deficits may have stemmed from his problematic processing in the distance channel.

Would he have similar deficits when moving his eyes? On the one hand, both eye movements and visually guided actions operate on visual information and in an eye-centered reference frame. Thus, these two processes may use the same representational format. On the other hand, Greenwood, Szinte, Sayim, and Cavanagh (2017) have shown that uncertainty in saccade landing positions is not correlated to uncertainty in visual perception, suggesting that different formats may be used even if position information is defined in the same eye-centered reference frames. Thus, we set out to explore which representational format the eye movement system may use.

Using a similar experimental setup, we asked MDK to saccade to a peripheral position, which was denoted by a single dot. Using an eye tracker, we recorded not only his saccade destinations in response to peripheral positions, but also trajectories during his saccade execution. We combined these two pieces of information to study the possible format of position information in MDK's eye movements.

Method

Apparatus

Stimuli were presented on a monitor screen that was viewed by MDK at a 40 cm distance. The screen extended 68.04° x 37.80° and refreshed at 60 Hz. A chinrest was used to stabilize MDK's head and to ensure that MDK directly faced the screen. We used an SR Research Eyelink 1000 Tower Mount eye tracker to record MDK's eye movements; it recorded at a 500 Hz sampling rate. A five-point calibration and a validation procedure were run before each testing block. This procedure was repeated until we could accurately track MDK's eyes with a mean error of less than 0.5°. The eye tracker was controlled by in-house MATLAB scripts.

Stimuli

Stimuli were single dots of 1° in size. In each trial, a dot could appear 6° or 8° away from the fixation, in one of eight directions: four principal directions on horizontal and vertical meridians, and four diagonal directions.

Procedure

Each trial started with a fixation mark. When MDK successfully maintained his fixation on the mark and reported that he was ready, a stimulus dot flashed for 200 ms or stayed on the screen until MDK finished saccading on the target. MDK was allowed to saccade to the dot's position as soon as he saw the dot. Each trial ended when MDK fixated on a position on the screen for three seconds without saccading to another position. This position was denoted as his saccade response to what was presented earlier.

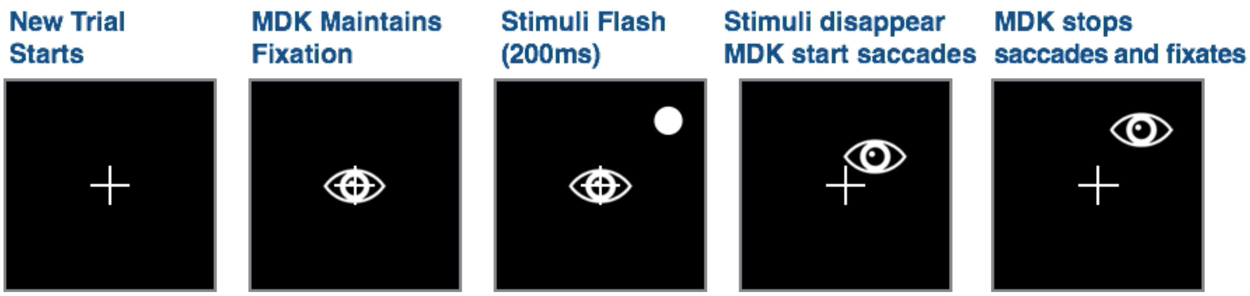


Figure 2-9 MDK saccades to a peripheral position

Results and Discussion

We first analyzed the final landing positions of MDK's saccades when he responded to different target positions. Figure 2.10 illustrates his landing positions in both the stay-on condition and the flash condition. When the stimulus stayed on the screen, MDK eventually saccaded to correct positions. When the stimulus only flashed for 200 ms, MDK saccaded to a peripheral position, then stopped. In this condition, he never reached the true distance, but preserved the angular bearings relatively well, similar to his fixation bias responses in the dot-pointing task. The results of his landing points favor a polar coordinate format in his representation.

MDK never reached a landing point with a single saccade. Figure 2.11 illustrates several first saccades in a condition when the stimulus stayed on the screen. These trajectories show that MDK never saccaded to the target destination in his first saccade. Therefore, the structure of the landing point may only reflect the structure of positions in MDK's visual memory. How position is encoded in his execution of each eye movement still requires an analysis of his saccade trajectories.

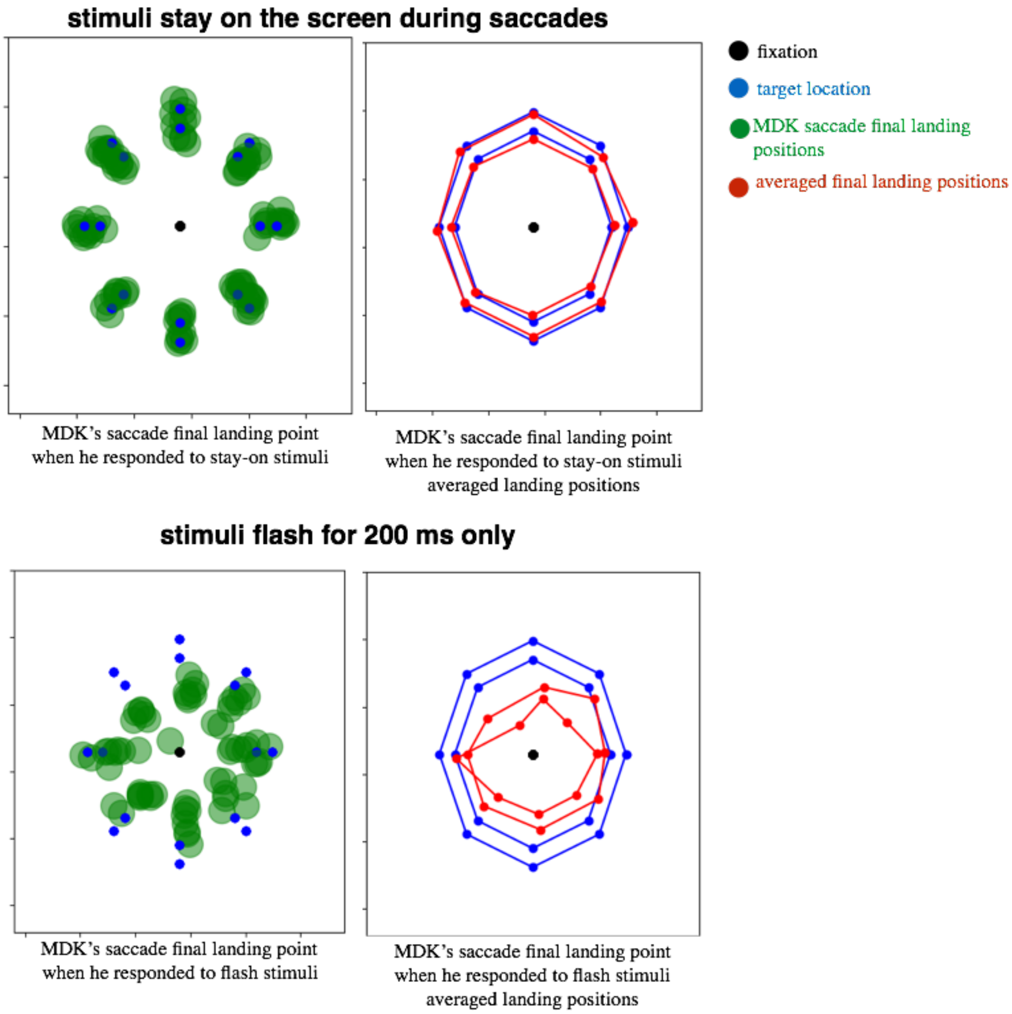


Figure 2-10 Landing positions of MDK's saccade to peripheral stimuli in stay-on and flash conditions

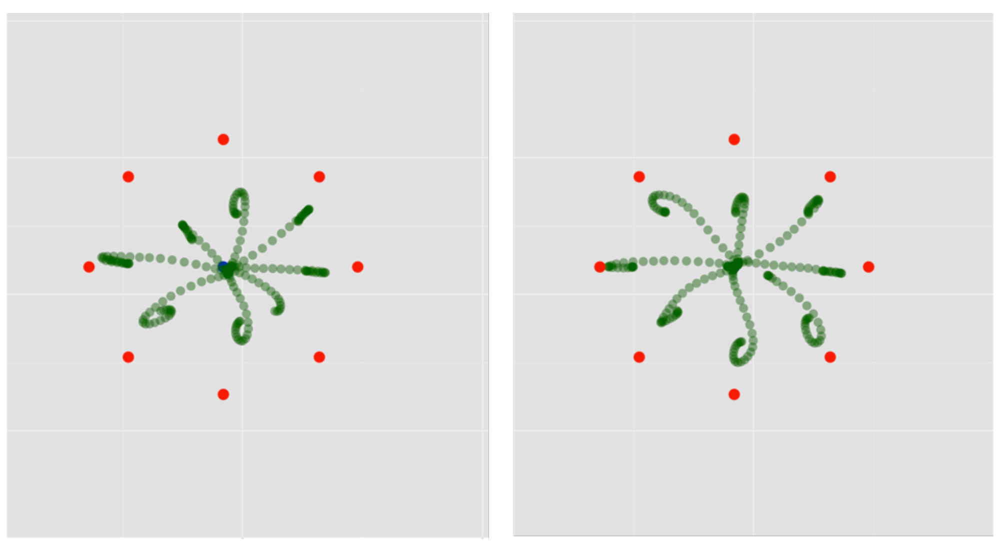


Figure 2-11 Samples of MDK's first saccades to stimuli that stayed on the screen

Thus, we further analyzed MDK's saccade trajectories. We decomposed his saccades into a horizontal and a vertical subcomponent. If MDK's saccades, especially his oblique saccades, were planned in a polar coordinate system, the dynamics of horizontal and vertical subcomponents should be synchronized. That is, after a period of fixation, two subcomponents should start to move at the same time and reach the peak speed at the same time. Hence, we calculated the starting latency and peak latency in each subcomponent in each oblique saccade. We defined a saccade to be an oblique one if its moving direction was at least 15° away from the vertical or horizontal directions. We extracted 10ms data before the saccade as the baseline data. The starting latency was determined by the first time a subcomponent's speed raised from the baseline speed and crossed a threshold speed. Here, we set the threshold to be 15°/s. The peak latency was determined by the time when a subcomponent reached its maximum speed. MDK's maximum speed could reach 185°/s when he executed a 4° saccade. Figure 2.12 shows that the starting latencies of the two subcomponents were not on the same latency line. Sometimes the vertical component started earlier, while other times the horizontal component did. The two subcomponents did not have the same peak latency either. In many cases, the horizontal subcomponent reached its peak speed before the vertical subcomponent did. The results showed that the two subcomponents of MDK's oblique saccades were rarely synchronized, suggesting that the oblique saccades might not be planned by a single signal that directs the saccade direction. Instead, to saccade to the target position, the destination position might be represented as a vertical signal and a horizontal signal that are separately planned and executed by two channels in the eye movement system. In other words, the asynchrony of subcomponents of oblique saccades disfavors a polar format in the position representation in the eye movement system; instead, the system may represent positions in terms of magnitude of horizontal and vertical translations, similar to a Cartesian format.

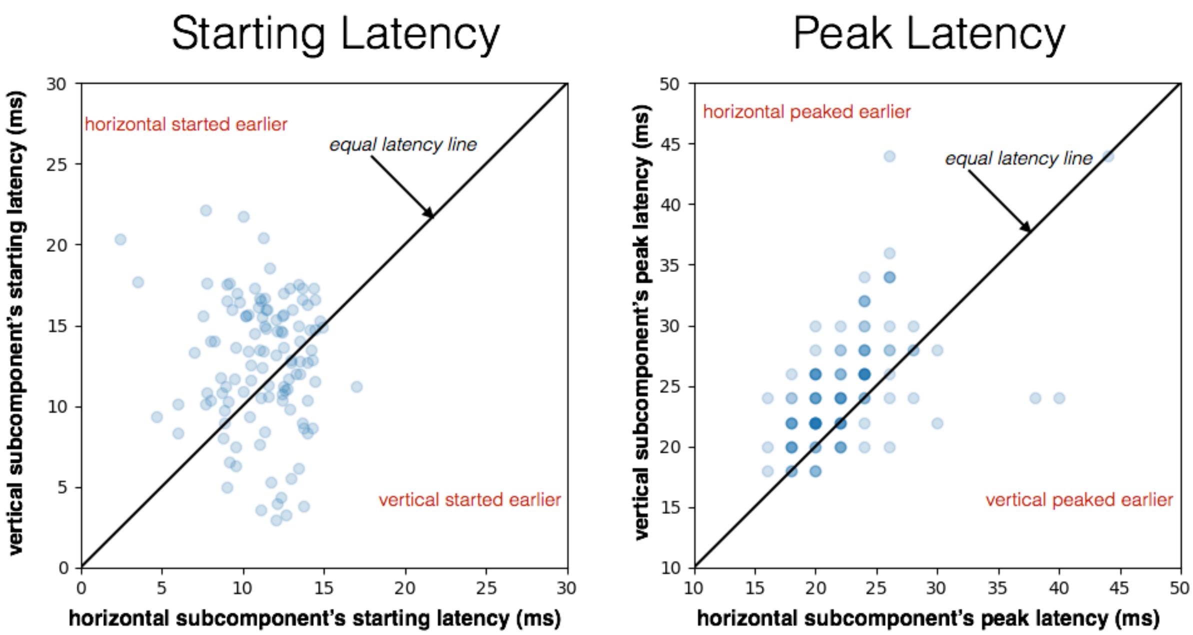


Figure 2-12 Starting latencies and peak latencies of MDK's vertical and horizontal subcomponents in his oblique saccades

General Discussion

We have tested MDK for more than two years. During this period, we have tried many different tasks to understand the nature of his spatial deficit. Our explorations have been constrained by limited variations of tasks in which MDK can understand and follow our instructions and can provide steady and easy-to-understand responses. In our results, we clearly observed that his visually guided action was executed in an eye-centered reference frame. The key to maintaining such a reference frame of position was to keep a visual-action feedback loop. We also found his visual representations define positions both in an egocentric reference frame and in an allocentric, object-centered one. Integration of these different frames depended on the task requirements of his interaction with the physical world. Furthermore, we also studied MDK's eye movements and found that the vertical and horizontal subcomponents of his oblique saccades were rarely synchronized. Such asynchrony suggests that the vertical and horizontal displacement information was separately encoded in the eye movement system, similar to a Cartesian format. Our results are consistent with previous saccade research (Bahill & Stark, 1975; 1977).

Box: Summary of MDK's testing results and implications	
Results	**Implications**
2.1.1: MDK's visually guided pointing responses were biased towards his fixation position.	Visually guided actions used an eye-centered reference frame.
2.1.1: MDK's pointing responses failed to keep a similar structure when his eyes were open or closed. When he closed his eyes, MDK's pointing shifted downwards overall.	Visually guided actions needed online visual feedback to maintain the eye-centered reference frame; otherwise, actions may have used positions defined by other parts of the body, such as a shoulder-centered reference frame.
2.1.2a: Unimanual pointing to endpoints of bars was less foveally biased than to dots of matched distance.	MDK may have engaged information about a bar, possibly its length, to help him point to the peripheral position on that bar.
2.1.2b: Bimanual responses to endpoints of bars were qualitatively different from a combination of two unimanual responses in the same condition.	MDK may have further engaged information about the length of a bar when he needed to reach both endpoints of the bar. In this case, an independent, object-centered representation of space was engaged to help describe peripheral positions.
2.1.2c: Bimanual responses to endpoints of bars were different from bimanual responses to two dots in the same positions.	Results in 2.1.2b cannot be simply explained by a bimanual coordination mechanism, further suggesting that MDK used object information when he pointed to both endpoints of bars.
2.2: Horizontal and vertical subcomponents of MDK's oblique saccades were rarely synchronized.	The eye movement system did not encode destination positions in terms of direction and angles as in a polar system. It may have separately encoded vertical and horizontal displacements to the destination position.

Table 2.1 Summary of MDK's testing results and implications

Chapter 3 Variance and Bias in Visual Position Representations

Inspired by our testing of MDK, we hypothesize that the human visual system uses a coordinate system to describe positions in space. In theory, a coordinate system has the power to support all possible algebraic computations that are necessary for visual behaviors. Empirically, previous evidence from neuroscience and psychological studies also supports this hypothesis. Single-neuron recording studies have shown that neurons in the parietal cortex are selectively activated by azimuth and elevation angles (Anderson, 1995).

Our goal is to understand what coordinate system does the visual system use. To this end, we developed two hypotheses: that it uses a Cartesian system, and that it uses a polar system. To distinguish between these two, we measured uncertainty in position processing and then characterized their noise distributions. We reason that different coordinate systems predict different noise distributions: noise may stem from horizontal and vertical displacements from cardinal axes in a Cartesian system, or it may stem from radial distances and angles in a polar system. Different sources of noise can result in different noise distributions in space. Two confusing positions in one representation may be clearly separable in another representation. Thus, we focused on studying the noise in people's responses.

Specifically, this chapter is divided into two sections. In the first section, we focus on the shape of the noise distribution. We try to understand when our visual system, or other cognitive system that uses visual information, represents positions as Cartesian / polar coordinates. Then, in the second section, we look at biases when people respond to a peripheral target. Previous research has shown that we perceive a position to be closer to our fixation than it should be. In this section, we consider that bias comes from an optimization process under different representations.

Position Variances and Representational Format

Experiment 3.1 Noisy Representations of Peripheral Positions

In this first experiment, we set a foundation for the rest of the study. Our goal was to understand the noise in position representations. To reveal the format of such representations, we designed a simple task that required people to repeatedly discriminate a peripheral target position or respond to it while maintaining fixation.

In the perception task, we presented a target stimulus and a comparator stimulus successively and asked people to determine whether or not the two stimuli appeared at the same position. In a noisy representation of the target, people should make more errors when comparators appear at a position where the target is highly probable given a noisy memory. In contrast, when comparators appear at a position where the target is not likely to be, people should easily tell the stimuli apart. We identified the positions where people confused a target and a probe, and we then constructed a probabilistic model of the representation's noise distribution. With such a distribution, we could further characterize the variances.

In the motor response task, we asked people to saccade to a target stimulus in the periphery, or to point to a target stimulus in the periphery while maintaining fixation. We recorded people's saccade landing positions or pointing positions on a touch-screen. We asked people to respond to the same target position multiple times so that we could collect a set of response positions given that target position. We could similarly construct a probabilistic model of response distributions and characterize the variance of each distribution.

A Cartesian model and a polar model predict different shapes for response noise distributions, as illustrated in Figure 3.1. A Cartesian noise distribution has the shape of a circle or an ellipse. Positions on the left and right of the target or above and below it are equally confusable. In contrast, a polar noise distribution should have the shape of a teardrop. It is anisotropic and naturally oriented towards the fixation, which is assumed as the origin of an eye-centered coordinate system. Thus, we could test which model better explained responses.

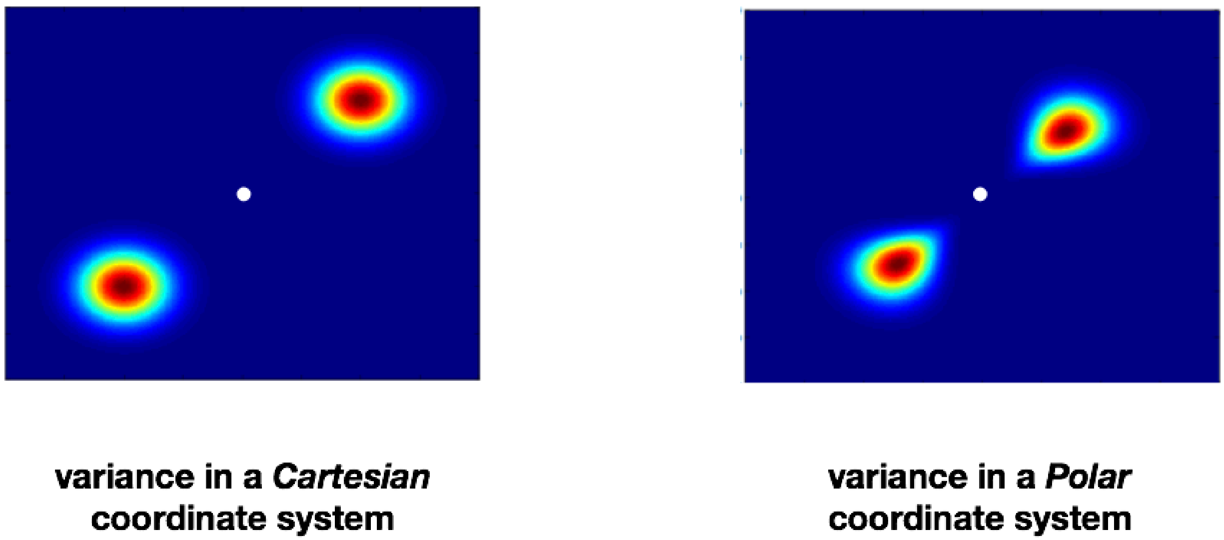


Figure 3.1Model prediction of shape of variances in position distribution

Method

Participants

Forty-one Johns Hopkins University undergraduates participated for course credit. All had normal or corrected-to-normal visual acuity. Participants were assigned two groups: a *perception* group (n = 21) and an *action* group (n = 20).

Apparatus

Stimuli were presented on a Macintosh iMAC computer with a refresh rate of 60Hz. The viewing distance was approximately 60 cm so that the display subtended 39.43° x 24.76° of visual angle. In the saccade task, the right eye of each participant was tracked using an SR Research Eyelink 1000 Tower Mount eye tracker sampling at 500 Hz. A five-point calibration and validation were run before the experiment, with a mean error of no more than 0.5°. The eye-tracking process was controlled by in-house MATLAB scripts.

Stimuli

Stimuli were generated using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). A *fixed* set of positions was used as target positions for both perception and motor response tasks. This set consisted of 16 positions at eight angular bearings (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 335°) with two eccentricities (6°and 12° visual angle). An additional set of 36 positions was randomly chosen before the experiment for the action group. This *random* set was used in a saccade task and a touch-screen pointing task as target positions.

Design and Procedure

Participants in the *perception* group were asked to discriminate peripheral positions. In the perception experiment, each trial started with a fixation dot that turned to red for 200 ms. Then, two Gabor patches, each of which subtended 1.5° x 1.5° of visual angle, appeared successively in the peripheral visual field. The first Gabor was presented for 200 ms. Then, after a 500 to 700 ms blank, the second Gabor appeared and stayed on the screen until a response was recorded. One of these two was a target and the other a comparator; their presentation order was counterbalanced. The target appeared at one of 16 fixed positions, whereas the comparator randomly appeared at a position near the target: in 80% of trials, the comparator's center was uniformly distributed in a 3.4° x 3.4° box around the target's center, while in the remaining 20% of trials, the comparator and the target appeared at the same position. Participants were required to maintain fixation throughout the experiment and to only use peripheral vision to discriminate stimulus positions. They responded whether the two Gabor patches were presented at the same position or not. For each target position, participants discriminated 50 comparator positions against the target position.

Participants in the *action* group were asked to saccade to and point to peripheral positions in two experimental sessions. In these action experiments, after the fixation dot turned to red for 200 ms, a Gabor patch flashed at a peripheral position on the screen for 200 ms, followed by an auditory cue that directed participants to respond. The target position could either be one of 16 fixed positions, which were identical to the positions in the abovementioned perception task, or one of 36 random positions in the random set. Different experimental sessions required different action responses. In the *touch-screen* session, participants reached towards and touched the perceived position on a touch-screen. During the response, they maintained their fixation. In the *saccade* session, participants saccaded to the perceived target position and fixated there for 400 ms to confirm their response. For each target position, the participant provided 25 responses in the touch-screen session, and 15 responses in the saccade session.

Before each experiment, a short demonstration allowed each participant to become familiarized with the instructions and procedures.

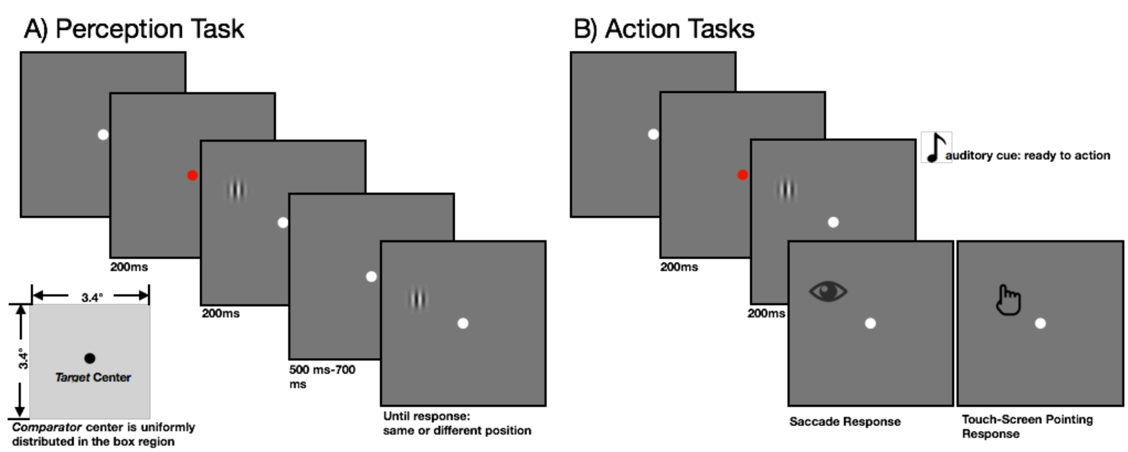


Figure 3.2 Experimental procedures for perception and action tasks in Experiment 3.1

Results

Our analysis focused on understanding the variances in noisy response distributions, and on determining which model better explained the shape of the noise distribution. To test the

Cartesian system hypothesis against the polar system hypothesis, we fitted the data with two bivariate Gaussian models with independent variances. In a bivariate Cartesian model, the variances come from the horizontal and vertical displacements, i.e. independent noise in each of the channels. Thus, we expressed the noisy position distribution as

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}\right]$$

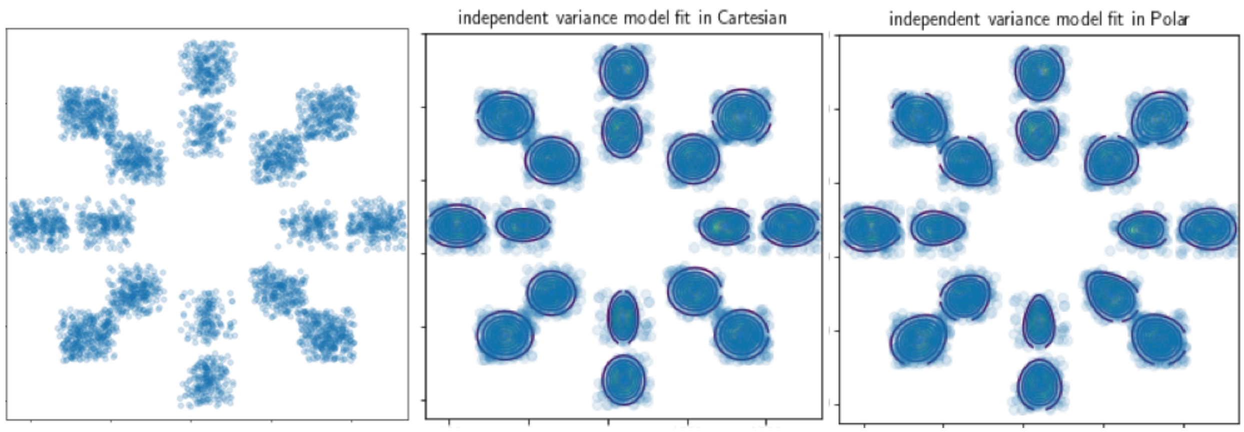
In contrast, in a bivariate polar model, the variances come from noise in the radial distance channel and the angle channel. Thus, we expressed the noisy position distribution as

$$\begin{bmatrix} \rho \\ \theta \end{bmatrix} \sim N\left[\begin{pmatrix} \mu_\rho \\ \mu_\theta \end{pmatrix}, \quad \begin{pmatrix} \sigma_\rho^2 & 0 \\ 0 & \sigma_\theta^2 \end{pmatrix}\right]$$

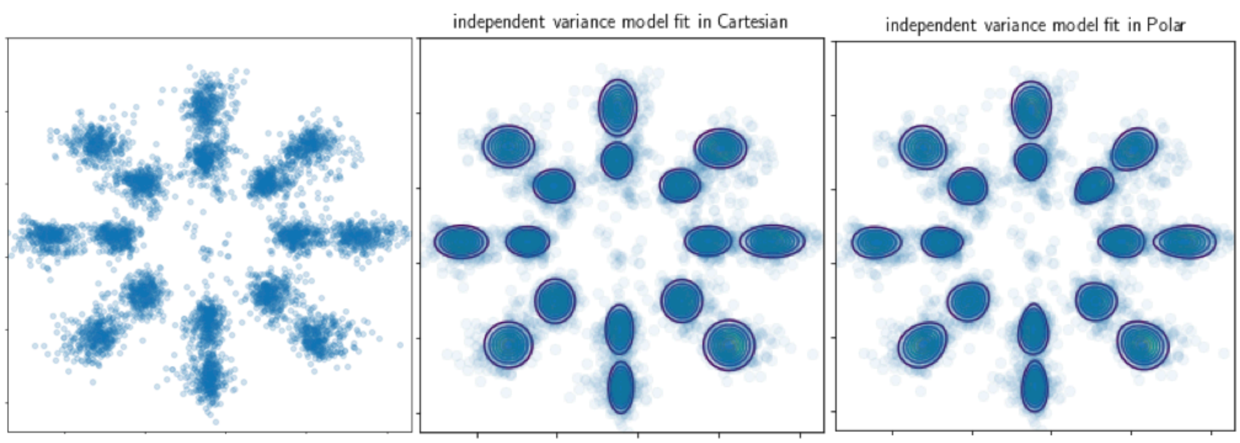
We used maximum likelihood estimations to fit both models to the data in the three tasks, and we compared which model fitted each data set better. We used the Bayesian information criterion (BIC) for model comparison. This criterion penalizes the complexity of a model, that is, the number of parameters given the number of observations. A model with a lower BIC score is regarded as a better model that describes the observed data well, and a difference of more than 10 in BIC scores is strong evidence that one model is better than the other (Raftery, 1995).

Raw Responses and Model Fitting Results

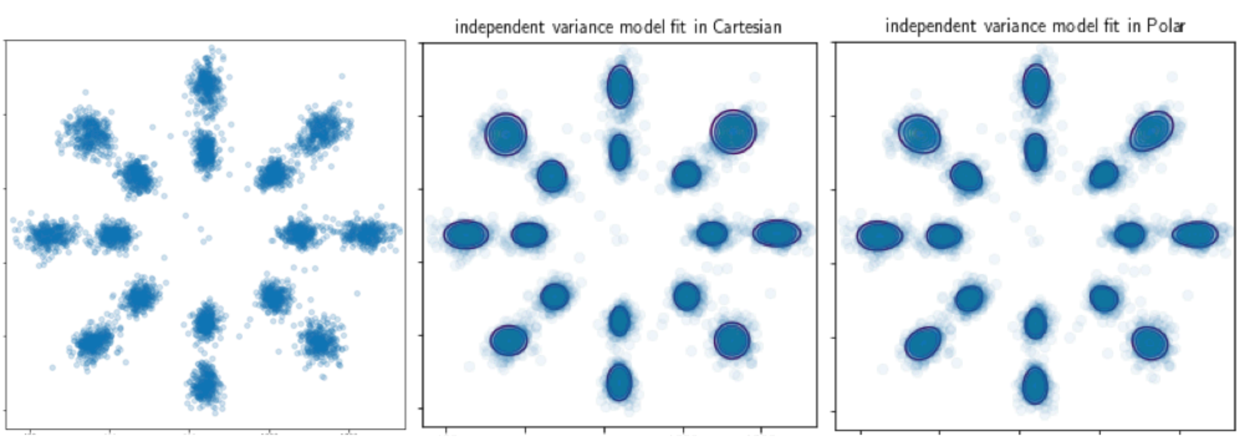
The responses to different target positions in each task are illustrated here. Contours in figures represent fitted distributions. Different models captured different proportions of data.



Perception judgment and model fitting distributions



Saccade responses and model fitting distributions



Pointing responses and model fitting distributions

Figure 3.3 Behavior responses and model fitting distributions from three tasks in Experiment 3.1

Model Comparison Results

To quantify which model was a better description for each task, we calculated BIC scores for each model in each task.

Data Set	BIC scores		BIC score differences	Preferred Model
	Cartesian model	**Polar model**	**Cartesian – polar**	
Perception: fixed set	34962.72	34863.44	99.28	polar
Saccade: fixed set	44716.34	44727.05	-10.71	Cartesian
Saccade: random set	54077.00	54187.89	-110.89	Cartesian
Pointing: fixed set	39150.94	39036.80	114.14	polar
Pointing: random set	71925.42	71844.58	80.84	polar

Table 3.1 Model fitting and model selection results in three tasks in Experiment 3.1

The results favored the bivariate polar model for the perception and visually guided action responses. In contrast, the bivariate Cartesian model fitted the saccade data better. A leave-one-participant-out cross-validation procedure produces similar results.

In addition to the bivariate Gaussian model with independent variances, we tested two other Cartesian models: a same-variance Gaussian model and a correlated-variance Gaussian model. In the first one, the variances are always the same for horizontal and vertical displacements; and in the second, the variances of the horizontal and vertical displacements are correlated. These additional Cartesian models yielded similar results as the aforementioned Cartesian model. The model selection results remained the same when these Cartesian models were compared to the polar model.

Discussion

Different formats of position representation should have different variances owing to having different sources of noise. Accordingly, models of these formats predict different shapes of noise distributions in position responses. We fitted behavioral data with a Cartesian and a polar model to study the sources of variance in position knowledge, and we used model selection methods to identify the model that described the observed data better. Our results indicate that perception and visually guided pointing are best characterized by a polar coordinate system, whereas saccade responses are best described by a Cartesian coordinate system. These results suggest that perception, saccade, and visually guided action systems use different underlying coordinate systems to represent peripheral positions.

That visual perception and pointing utilize a polar model was implied by our previous neuropsychological observations in MDK's testing. However, the present behavioral study extended our observations to a formal experiment and confirmed the polar hypothesis with an abundant amount of data. The polar hypothesis is consistent with previous electrophysiological results that neurons in visual cortices are sensitive to angle and distance (Andersen, 1995). Furthermore, this polar hypothesis describes a structure of space that readily explains an observation that variance increases as target positions become further away from the fixation, as shown in Figure 3.3. A polar coordinate system relies on its origin point to describe a position. This description carves the space in unit grids of various sizes: the further away from the origin, the larger the size of a unit grid, as illustrated in Figure 3.4. A smaller unit grid near the origin more finely describes areas near the origin than areas that are far away from the origin. Thus, variance of responses is smaller when the position is near the origin, and becomes larger when the target position is far in the periphery. In an eye-centered reference frame, the origin point is the fixation. Thus, we observe that responses to positions far in the periphery have larger variance. In contrast, a Cartesian coordinate system carves the space in unit grids of equal size regardless of a region's distance from the origin. The Cartesian coordinate system cannot naturally explain a larger variance when target positions are farther in the periphery.

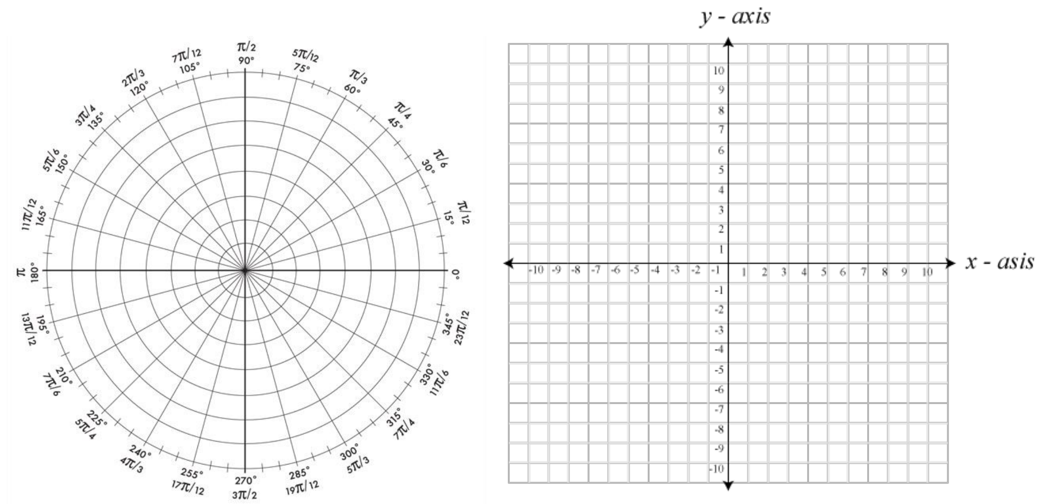


Figure 3.4 A polar system carves space in unit grids of different sizes; a Cartesian system carves space in unit grids of equal size

Position representation in saccade is different from that in visual perception. This result is consistent with our analysis of MDK's saccade trajectories. The reason may be biological constraints on saccade execution: humans have two pairs of muscles that control the eyeball's vertical movements, and one pair of muscles that control horizontal movements. Previously, King, Lisberger, and Fuchs (1986) studied human and monkey saccades and proposed that the saccade system has a single source of direction and distance in position representations. Information from this source is then calculated and sent to different muscles, which is then translated into Cartesian commands. Therefore, the authors propose that the position representation in the saccade system is a polar coordinate system. In contrast, Bahill and Stark (1975, 1977) showed that the horizontal and vertical subcomponents of an oblique saccade do not always synchronize. Instead, each subcomponent seems to have independent information sources, suggesting a Cartesian model in saccade execution. Grossman and Robinson (1988) improved this model with a coupling component, and provided further evidence of a Cartesian model in saccade. In line with previous studies, our computational analysis of human saccade landing positions also favors the Cartesian model over the polar model for saccade position representation.

Recently, Greenwood, Szinte, Sayim, and Cavanagh (2017) have shown that saccade landing positions are not correlated with perception uncertainty. The authors suggest a distinction between spatial representations in these two systems. Here, we clarify this distinction and posit that the differences between these two systems come from the different formats used in their position representations.

Central Tendency and Representational Format

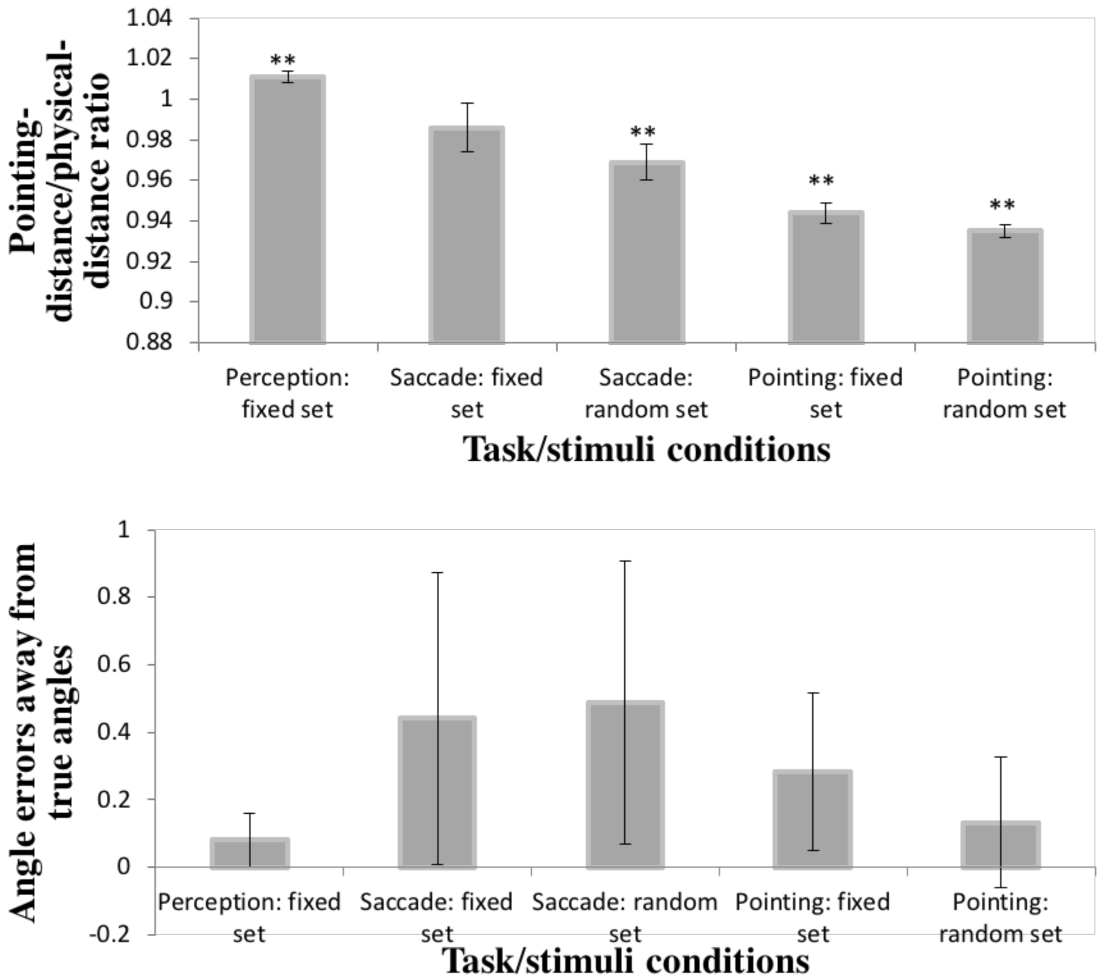
In addition to variance, the central tendency of response distributions may also inform us about the format of the underlying coordinate system. Previous research has shown that people tend to mis-localize peripheral positions in perception, saccade, and pointing. People often perceive, saccade, and point to a peripheral position that is closer to their fixation than a true target position (Tsal and Bareket, 2005; Sheth & Shimojo, 2001, Ross, Morrone, & Burr, 1997). In this section, we first examine where the average response position is. If there is bias, can a coordinate representation explain it?

Experiment 3.2 Fixation Bias Revisited

We first revisit data from Experiment 1 here, but focus on the central tendencies in these noise distributions, that is, the averaged positions of the responses. Our early analysis suggested different position representation formats for perception, saccade, and pointing. In this section, we examine whether averaged positions in these data also show different characteristics.

Results and Discussion

We compute the ratio between the pointing distance and the true stimulus distance from the fixation as a measure of bias; furthermore, we also compute the difference between the pointing angle and the true angle as a measure of any bias in direction. We expect a small deviation of pointing angle to the true angle if a polar coordinate system is used, because angle is a primitive in a polar representation that does not require further computations.



*p < 0.05, ** p<0.01, Error bars denote SEMs

Figure 3.5 Mean pointing distance/physical distance ratio in different tasks and mean angle errors

First, angles in all responses are not significantly different from true angles. We thus focus on understanding whether responses preserve the accurate distance of the target locations. In our results, perception responses do not have biases towards fixation. In fact, the mean position in perception is slightly further away from the fixation than the true position ($p < 0.001$). Previous studies observed a similar localization bias towards the periphery when letters were randomly placed in peripheral regions, even when attention was directed towards that region (Tsal and Bareket, 1999), but the magnitude of perceptual bias in the present results is about 1% of the real distance. This small bias in perceptual tasks suggests that people have relatively accurate, though not necessarily precise, perceptual representations of peripheral positions. This accurate representation, then, may not be the reason for the fixation biases observed in tasks involving an action, including saccade and pointing. In contrast, saccades and pointing responses show biases towards the fixation. When the set of locations is fixed, saccade responses are not significantly different from the true target locations. When the set of locations is random, saccade responses are biased towards the fixation direction, with a magnitude of 3% of the true distance. Pointing responses have larger biases towards the fixation. When the set of locations is fixed, the magnitude of bias in pointing responses is about 5.5% of true distance; when the set of locations is random, the magnitude of bias reaches 6.5% of true distance. The results show that knowledge of a stimulus's possible location can help to decrease response biases. However, the general bias towards the fixation in both action tasks requires further explanation.

Why are action responses biased towards the fixation? In previous research, it was hypothesized that short-term memory averages the position of the target and the position of fixation and leads to this foveal bias phenomenon (Sheth & Shimojo, 2001). Here, we address this issue from the perspective of perceptual decision-making, and hypothesize that localizing a peripheral position involves an optimization process that minimizes expected cost from noise in perception and action. We describe the computational process as follows.

When we respond to a peripheral position, we want to make an accurate response to effectively interact with objects there. However, our perception and action are noisy. When we respond, we need to take into account these uncertainties and make an optimal decision that minimizes the cost of errors due to noise. This requires an optimization process that minimizes expected costs of a decision. Equally probable errors may have different costs. If the probabilistic distribution is in a polar coordinate system, two equally probable error positions from the same distribution bear different amounts of cost because the polar system carves the space in unequal unit grids, as shown in Figure 3.4. A position that is away in the periphery bears a larger cost because it actually has a larger deviation from the true position in space. The other position that is close to the origin bears a smaller cost because it actually deviates less from the true position in space. Therefore, under the assumption that positions are represented in polar coordinates, in this optimization process it is optimal to respond to a position closer to the fixation given the uncertainty in perception and action. In contrast, if the probabilistic distribution is in a Cartesian coordinate system, two equally probable error positions in the distribution bear the same amount of cost because the Cartesian system carves the space in equal-sized unit grids. Therefore, in the optimization process, it is optimal to respond close to the true position given perception and action uncertainties.

Our results show that pointing responses are clearly biased towards the fixation. Considering that the model selection results in Experiment 3.1 indicate that both visual perception and visually guided actions use a polar coordinate system to represent positions in space, this decision bias is consistent with the prediction of the optimization hypothesis that uses a polar model representation. In comparison, saccade responses have smaller biases. This is probably because the Cartesian coordinate representation in the saccade system, as suggested in the previous model selection results, reduces biases in the optimization computation with its equal-sized unit-grid structure of space.

Experiment 3.3 Telephone Game and Accumulative Fixation Bias

In our second attempt to use central tendency of noisy position distributions to investigate the underlying format of the visual coordinate system, we conducted an experiment that accumulated biases from central tendency and thus amplified the differences between the two models' predictions.

In this experiment, participants moved a mouse to point to a peripheral target stimulus. Unbeknownst to them, a target stimulus in future trials was determined by their current response positions. This process is similar to a telephone game, where people pass on information based on what they hear from others. In this experiment, participants played a telephone game that communicated peripheral locations with a computer program. In the initial iteration of a game, the computer program briefly presented a stimulus at a peripheral position on the screen. A participant perceived the stimulus and clicked on the screen to indicate where s/he thought the stimulus was presented. In the next iteration, then, the computer presented a new stimulus at the same position where the participant had last clicked on the screen. As the game continued, every time the participant clicked a position on the screen, the computer presented a stimulus at that same position in the next iteration. The game was played for multiple iterations.

If a perceptual bias is constant and systematic, it should exist in each iteration of the game, and accumulate after multiple iterations. As a result, a trajectory of the bias accumulation process may amplify characteristics of the underlying representation. For example, 1) if a position representation follows a symmetric noise distribution around the true position, the trajectory should wander around the true position after several iterations; 2) if biases come from a single coordinate in a Cartesian coordinate system, the trajectory may mainly move horizontally or vertically towards principal axes; and 3) if biases reside in radial distances in a polar coordinate system and this coordinate system always has accurate direction information, the trajectory may move towards the fixation along the angular bearing from the starting point to the origin.

Method

Participants

Fifteen Johns Hopkins University undergraduates participated for course credit. All had normal or corrected-to-normal visual acuity.

Stimuli, Design, and Procedure

To study accumulative perceptual bias, we designed 16 independent testing chains. Each chain started at one of eight starting positions, so that each starting position had two testing chains. These eight positions were 10° away from the fixation, in one of eight directions (0°, 45°, 90°, 135°, 180°, 225°, 270°, and 335°).

Each trial randomly picked one testing chain and presented a stimulus accordingly. Each trial followed a similar procedure to that in the previous touch-screen pointing experiments, except that participants moved a mouse cursor to point to a peripheral position on the screen. Each trial started with a red fixation dot that lasted for 200 ms. Then, a Gabor patch (1.5° x 1.5° of visual angle) appeared at a peripheral position for 200 ms, followed by an auditory cue that directed the participant to be ready. After the auditory cue, participants pointed to the perceived position without time constraints.

Unbeknownst to participants, each experiment consisted of two parts without a pause between them. The first part was a *baseline* part, and the second a *telephone game* part. In the *baseline* part, each testing chain only presented a stimulus at the starting positions, and participants' responses to the starting position were collected. This part included 14 trials for each testing chain. In the *telephone game* part, each testing chain presented a stimulus at the position of the participant's early responses. After the participant pointed to a position in the current trial, the testing chain updated its stimulus position to be the participant's responded position, and presented a stimulus at this new position the next time. This part comprised 20 trials for each testing chain.

Before the experiment, participants were told to respond to peripheral positions as accurately as possible because their responses were going to be passed onto future participants, and the experiment was designed to test how accurately position information could be preserved across participants. After the experiment, we asked participants whether they had noticed that they were actually playing a game against themselves, a game in which their early responses determined their later experiment stimuli positions.

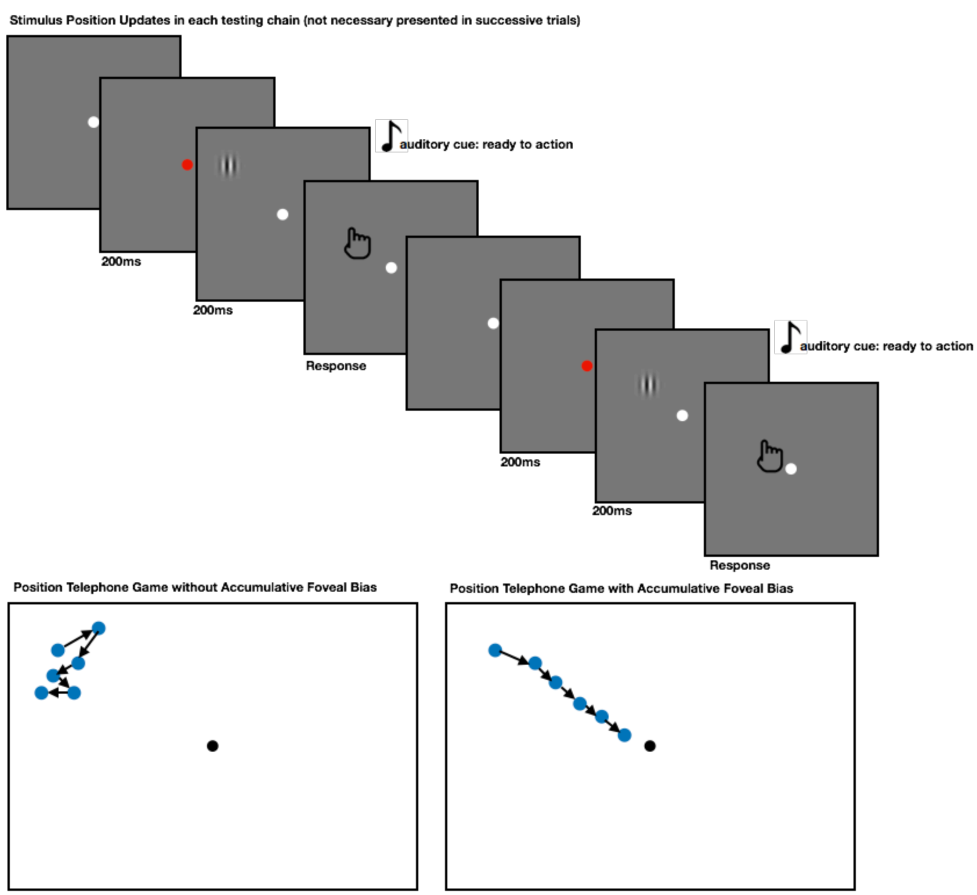


Figure 3.6 Procedures in Experiment 3.3 and model predictions of trajectories in the telephone game

Results

In the debriefing after the experiment, no participant reported knowing that he or she had played a telephone game against him- or herself. Despite being unaware of the telephone game, however, participants gradually pointed towards the fixation and usually ended in a position right next to the fixation. All responses were pooled together and are illustrated below, where we can clearly see eight arms to which most pointing responses converge. Point distances at each step in the telephone game are shown in the middle panel. It is clear that people pointed to roughly the same distribution of distances when we presented a target at fixed positions, and then quickly pointed towards the fixation in the majority of cases when the telephone game started. On the right, we show mean pointing distances at each game step for each testing chain. Clearly, after the telephone game started, average pointing positions quickly converged to the fixation position regardless of where the starting position was.

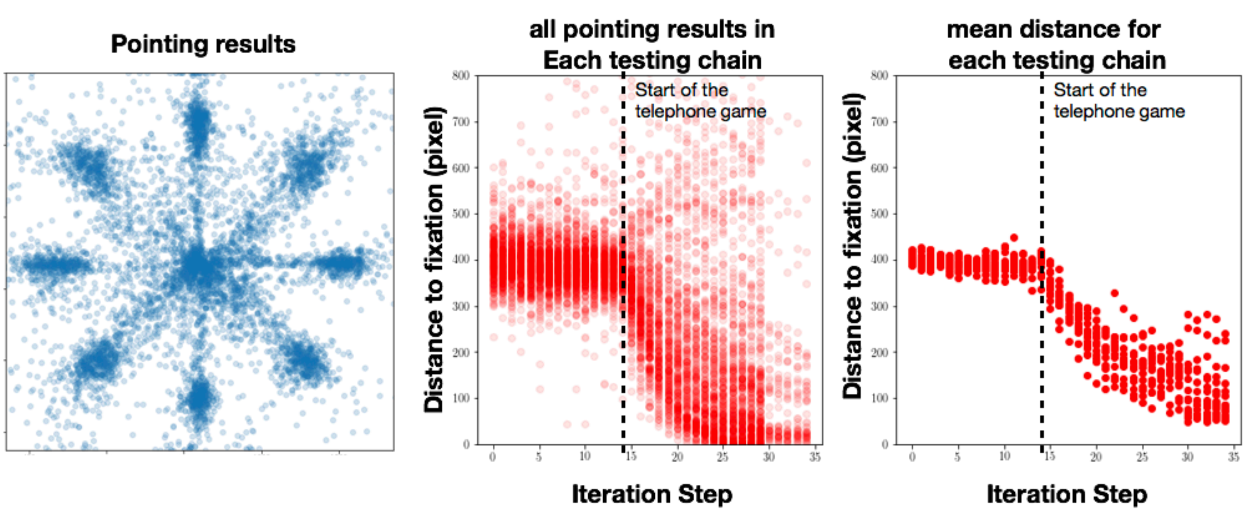


Figure 3.7 Pointing responses in Experiment 3.3. People gradually pointed to their fixation as they repeatedly pointed to their previously indicated positions. Average pointing distances from the fixation decreased as the iterative process continued.

To understand whether and how a representational format could affect the response patterns in this experiment, we simulated two computer participants, one with a Cartesian coordinate representation of positions (the Cartesian player) and the other with a polar coordinate representation of positions (the polar player), and asked them to play the same iterative telephone game. To make the Cartesian player and the polar player behave in a similar way to human participants, we used the model fitting results from Experiments 3.1 and 3.2 to characterize their position representations. Using the decision-making process described in Experiment 3.2, in each iteration of the game, the Cartesian player formed a Cartesian model probability distribution of the perceived position. To decide where to response, the Cartesian player drew a sample from the distribution and applied a random bias towards one or both cardinal axes. The magnitude of the bias was the same as the estimated bias magnitude in human participants during the *baseline* trials. The polar player formed a polar model probability distribution of the perceived position. To respond, the player drew a sample from the polar distribution and applied an underestimation bias to the distance of the sample in the actual response. The magnitude of the bias was the same as the one used by the Cartesian player. Similar to human participants, in each game, each simulated participant completed 14 *baseline* trials and 20 *telephone game* trials. Figure 3.7 illustrates the response distributions after 100 games for each stimulated participant.

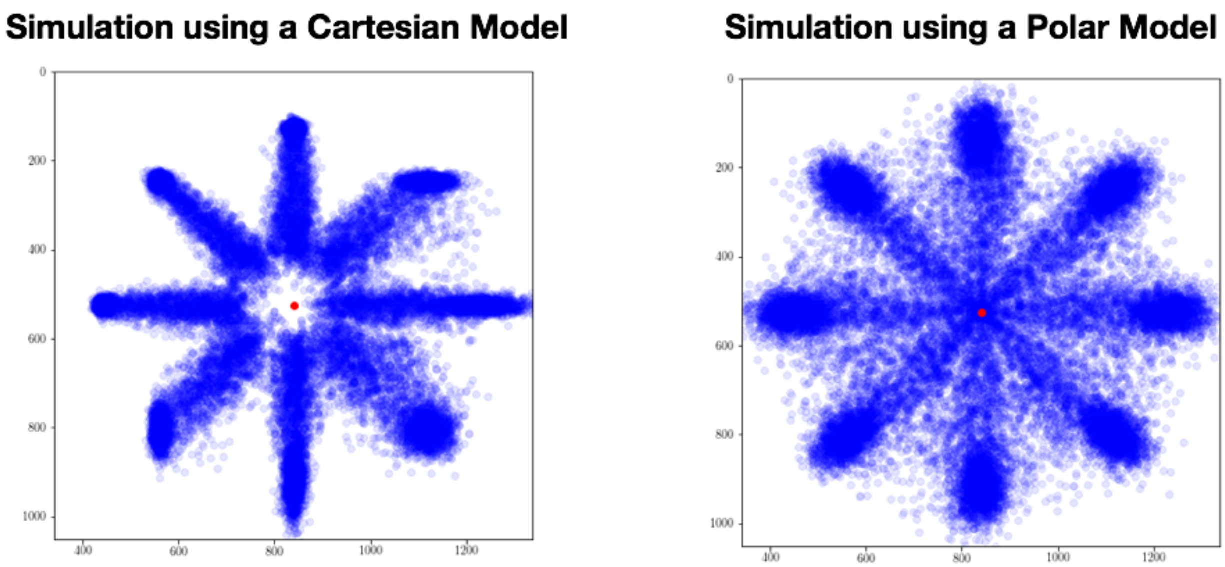


Figure 3.8 Model simulation of the iterative pointing process. A Cartesian model generates a new position each time based on the position of the last response. A bias coefficient is imposed on the generation process and the amplitude as well as on the distribution parameters.

Both simulated participants showed an accumulated response bias towards the fixation, but their overall response distributions were different. In the simulations, the two participants were modeled with the same magnitude of response bias in each iteration, but the magnitude of their accumulated biases after 20 telephone game trials differed. Figure 3.7 shows that the Cartesian player stopped at an area away from the fixation, whereas the polar player always stopped at the fixation position. That is, the polar player accumulated a larger foveal bias than the Cartesian player did while playing the same game. The polar player's response distribution and the magnitude of the accumulated bias were more similar to those in human responses (Figure 3.6). After 20 telephone game trials, human participants usually accumulated response biases so large that the game ended up presenting stimuli near the fixation. Thus, the polar player behaved similarly to human participants. It showed a large magnitude of accumulative distance bias in its responses. In contrast, the Cartesian player behaved less similarly to humans because it showed less accumulative distance bias. The behavior difference between the two simulated participants can only be attributed to the format of their position representations. Thus, that the polar player behaved similarly to human participants suggests that humans use a polar coordinate system to represent positions in space.

Discussion

In this experiment, we asked participants to respond recursively to positions where they had previously clicked on the screen, and used this procedure to accumulate response bias and variance. To further understand how this accumulative response bias reflects the underlying format of position representations, we simulated behaviors from two participants with different representational formats: one (the Cartesian player) used a Cartesian coordinate system and the other (the polar player) a polar coordinate system to represent positions. In Experiment 3.2, we addressed the pointing response as an optimization process that minimizes expected response costs under the uncertainty of perception and action. We also reasoned that a polar coordinate system naturally leads to a foveal bias in the optimization process because it carves the space with unequal-sized units, whereas a Cartesian coordinate system does not require a foveal bias in the optimization process because it carves the space with equal-sized units. If this optimization algorithm is applied repeatedly, as our current experimental procedure required, different coordinate systems predict different amounts of accumulative foveal bias by the end of the game. In this experiment, we even imposed a bias component on the Cartesian player and required it to elicit a response bias of the same magnitude as the polar player naturally did. After 100 simulations, we showed that the polar player accumulated a large foveal bias and ended up pointing at the position of fixation, whereas the Cartesian player accumulated a smaller foveal bias and never reached the fixation position at the end of the game. This simulation difference shows that response bias is strongly affected by representational format.

Bearing the stimulation results in mind, we can better understand human responses. When the human participants played the telephone game for the same number of iteration as the simulated participants did, the human participants accumulated a large response bias towards the fixation. They often ended up pointing at the fixation position by the end of the game, and the average distance of their pointing positions to the fixation steadily decreased as the game continued. Of the two simulated players, our results show that the humans behaved more similarly to the polar player. Because the initial magnitude of the bias in both simulated participants was the same as the initial bias in the human participants, our results suggest that the polar player is a better model to explain the human participants. Thus, the results of the human responses and the computer simulations together indicate that the underlying format of position representations in human visually guided action is a polar coordinate system. This result is also consistent with previous experimental results.

Compared with the experimental procedures used in previous studies, the telephone game provides a better opportunity to understand why people show foveal bias in their response and what representational format is used in computations. Using a single stimulus-response iteration, like in previous research, a Cartesian model and a polar model behave very similarly. A memory-averaged computation process and an optimization computation process can also provide a similar response. However, in the telephone game procedure, we can clearly differentiate between two hypothetical coordinate models by examining each model's accumulative responses. Moreover, the optimization process clearly specifies the hypothesis, describes computational steps, and derives detailed predictions in each iteration of the game. In contrast, it is unclear how a memory-averaged computation process can derive a similar level of algorithmic detail in explaining human behaviors in the game. Therefore, the results of this experiment not only suggest that visually guided action uses a polar coordinate system to represent positions, but also support the validity of the process of optimizing behavior output under uncertainty as an algorithm that leads to the foveal bias in human responses.

Chapter 4 Primitives of Position Representations in Perception

Different representational formats have different primitives. Primitives are basic units that are used to describe information in a representation. They make some aspects of information readily accessible and lead other aspects to be lost in the background. Recovering those hidden aspects of information requires computations that transform primitive information (Marr, 1982).

Two representations with different formats may describe the same content, but primitive information made explicit in one representation may be hidden in the other. Thus, these two representations can be useful for different computations. For example, a wave signal, such as a radio wave, can be represented in both the time domain and the frequency domain. A representation of this signal in the time domain explicitly describes how the signal oscillates over time, whereas a representation in the frequency domain explicitly describes how much signal falls in different frequency bands. Primitive information in the former representation, the temporal dynamic, is obscured in the latter representation; and primitive information in the latter representation, the loading on frequency bands, is difficult to access in the former. Therefore, computations of wave signals need to select a proper representation to process contents that are embedded in these signals. In general, representations and computations need to be matched appropriately in efficient information-processing systems.

In the case of space, specification of a coordinate system involves a selection of primitives. For example, a Cartesian coordinate system has primitives in the form of horizontal and vertical translations, whereas a polar coordinate system has primitives of radial distance and angular direction. Distance information in a polar coordinate system is hidden in a Cartesian coordinate system. To access or operate on distance information, algorithms that use a Cartesian coordinate system need to calculate distance, whereas those using a polar coordinate system can directly read out distance. Furthermore, extra computation may give rise to biases or noise. Thus, distance information in a Cartesian coordinate system may not be as reliable.

If we do not know the coordinate system in a representation, we can try to identify its primitives and then reveal the format. Specifically, this is our strategy here. We examine perceptual imprecision and biases in knowledge of position to infer the primitives used in vision. From these primitives, we may understand what underlying format is used in visual spatial representation.

In this chapter, we investigate spatial primitives by using discrimination tasks and position alignment tasks. In the first part, we test primitives in spatial representations directly. Inspired by the results in Chapter 3, we then test whether angular directions and distances are primitives in the visual processing of positions.

A polar coordinate system can directly read out angular directions of a position relative to the origin, but a Cartesian coordinate system needs to use trigonometric functions to transform horizontal and vertical displacements into angular direction, for example, $\theta = \arctan \frac{y}{x}$. In our study, we asked people to discriminate the angular bearing of two positions in space. If a polar coordinate system is used, we would expect that angular discrimination precision should only depend on the angular direction of a target location, and be independent of the radial distance of that target location. On the other hand, if a Cartesian coordinate system is used, we would expect angular discrimination to become increasingly imprecise as the displacement magnitude of the target location increases. Because displacement information itself becomes noisier and noisier when a position is farther away from cardinal axes, using this information to calculate the angular direction should introduce more noise.

Similar to angular direction information, a polar system can directly read out distance information, whereas a Cartesian system needs to compute distance from horizontal and vertical displacements. We also asked people to discriminate whether peripheral dots had the same distance from the fixation. If a polar system is used in visual perception, distance judgments should be accurate. If a Cartesian system is used, perceptual errors might appear in distance judgments.

In the second part, we tested spatial primitives in alignment tasks. Primitives in a Cartesian system can help to judge a position's translational displacement relative to other positions or main axes. These primitives can describe a straight line in space with a simple linear function, which should provide an accurate model that aligns dots on straight lines. Primitives in a polar system, in contrast, are bad at describing straight lines. A straight line requires a non-linear trigonometric model, for example, $r = \frac{1}{\sin(\theta) - \cos(\theta))}$, when angle and distance information is used. Accordingly, to align positions on a straight line, primitives in a polar systems may add noise in computations and lead to misalignment in perception. We therefore tested whether translations are primitives in visual perception by asking participants to align dots on a straight line. We extended the Vernier alignment tasks, in which people make fine judgments in the alignment of three dots. Using these tasks, we could examine geometrical calculations that align locations on a straight line and examine whether there are systematic perceptual biases in each alignment task.

Experiment 4.1 Comparing Angular Bearings

First, we tested whether information about angular bearing is primitive in perceptual computations. We asked people to compare angular bearings of two peripheral dots, and measured discrimination thresholds as perceptual precision. If angular information is primitive in representation, angular precision should be independent of the distance of the dots; otherwise, angular precision may decrease when dots appear further away in the periphery.

Method

Participants

A group of 18 Johns Hopkins University undergraduate and graduate students participated in the experiment in exchange for course credit. All participants reported normal or corrected-to-normal visual acuity and gave written informed consent. The experiment was approved by the Johns Hopkins University institutional review board (IRB).

Apparatus

All experiments took place in a dim, sound-attenuated room. There was no light source except for a computer monitor. All stimuli were presented on a Macintosh iMac computer with a refresh rate of 60 Hz. The viewing distance was 60 cm, so that the display subtended 39.43° × 24.76° of visual angle.

Stimulus and Procedure

Angular discrimination thresholds were measured with two-alternative forced choice (2AFC) staircase procedures, as illustrated in Figure 4.1. In each trial, two dots were successively presented on the screen. Each presentation lasted for 200 ms, with a 500 ms blank interval in between. The task was to judge whether the second dot rotated clockwise or counter clockwise from the first dot. One dot was a target dot, and the other was a comparator dot. The target dot was presented in one of nine conditions in a 3 by 3 experimental design: three angular bearings (45°, 180°, 270° from the polar axis) and three distances away from the fixation (6°, 8°, 10°). The comparators were the same distance away from the fixation as the target, but had different angular bearings. They rotated clockwise or counter clockwise from the target angle. The magnitude of the rotation was adjusted in a three-down-one-up staircase method, which resulted in a 79.4% convergence rate. The step size of the staircase was 0.1 log units. Each staircase consisted of four preliminary reversals and six experimental reversals. The geometric means of the experimental reversals were taken as the threshold for each staircase run.

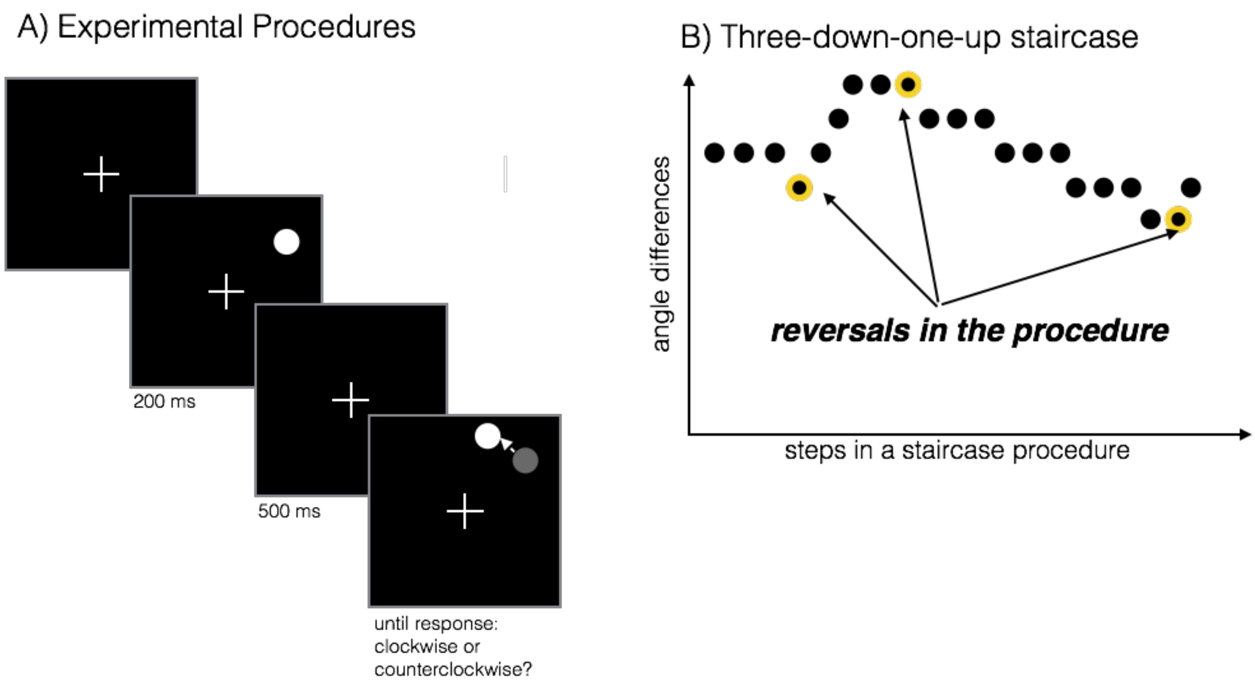


Figure 4.1 Schematic procedures of Experiment 4.1 and identification of reversals in the staircase testing procedure

Results and Discussion

The staircase procedures did not converge for four participants, thus these participants' data were excluded from further analysis. Figure 4.2 illustrates the magnitude of angular differences between a target and a comparator at each reversal point in each staircase procedure; data were averaged across 14 participants. The results showed that the last six experimental reversal points converged to a stable angular difference for each staircase procedure. There were no clear differences among the staircase trajectories of different target-comparator locations at different radial distances: trajectories at three different radial distances (6°, 8°, 10°) interleaved with each other.

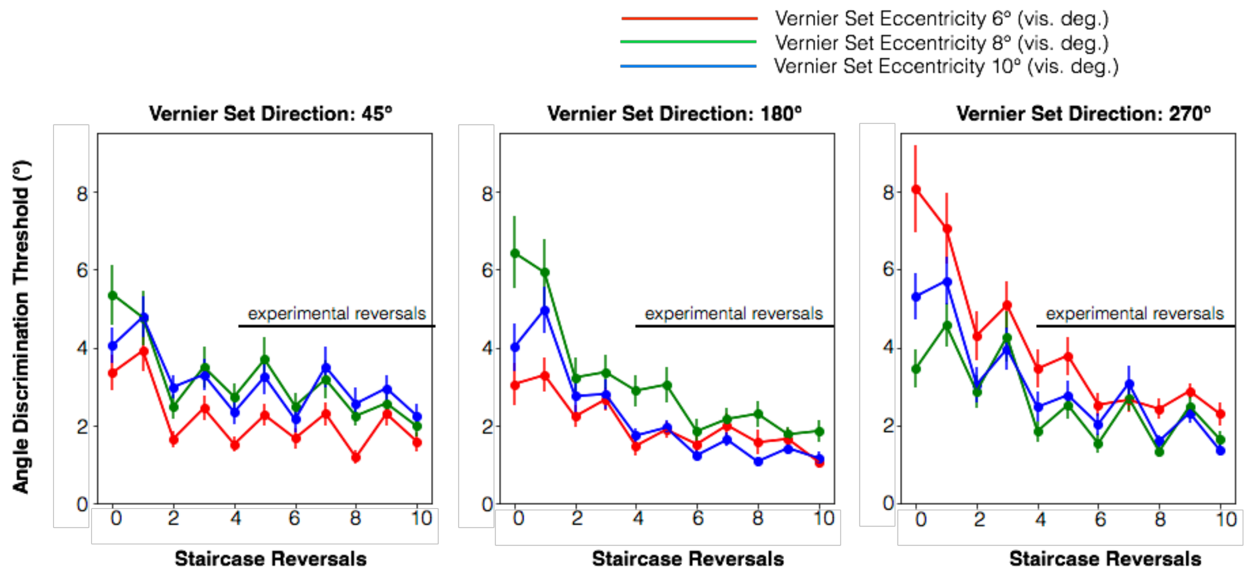


Figure 4.2 Group averaged reversal points in each staircase procedure

To confirm our observation, we estimated discrimination thresholds of 79.4% accuracy by calculating the geometric means of the last six experimental reversals. A discrimination threshold was estimated for each participant in each staircase procedure. We were mainly interested in the main effects of angular directions and distances. The former tested whether an angular perceptual variance was attributed to specific angular directions, while the latter tested whether perceptual variance changed as stimulus distance changed. Therefore, we conducted 3-by-3 repeated measures ANOVA with a simple additive model on the angular discrimination thresholds. The results showed a main effect in the target angular directions ($F(2, 26) = 3.44$, $p = 0.047$), suggesting that different target angles had different angular discrimination precision. There was no main effect in a target's radial distance ($F(2, 26) = 1.31$, $p = 0.286$). Therefore, our results showed no differences in angular discrimination precision in different radial distances in space, suggesting that our visual spatial representation directly uses angles as primitives in position descriptions.

Experiment 4.2 Comparing Distance

In the second experiment, we tested whether information about distance is primitive in perceptual computations. We asked people to compare the distance from the fixation to a peripheral target dot with the distance from the fixation to two other peripheral comparator dots. If distance information is primitive in representation, this distance comparison should be accurate. In contrast, a perceptual bias in comparison would indicate that distance is calculated from other primitives in position representations.

Method

Participants

Another group of 22 Johns Hopkins University undergraduate and graduate students participated in exchange for course credit. All participants reported normal or corrected-to-normal visual acuity and gave written informed consent. The experiment was approved by the Johns Hopkins University institutional review board (IRB).

Stimulus

A set of three dots was used, each of which measured 0.5° in diameter, as shown in Figure 4.3. The two comparator dots were placed at 8° away from the fixation and spanned a 45° angle from the fixation. The target dot was placed in between, but shifted along a radial line from the fixation. The shifting distance was from -0.875° to 0.875°, with a step size of 0.25°. A shifting distance equal to 0° meant that the target dot was at the same distance away from the fixation as the comparator dots were. When the shifting distance was negative, the center dot was closer to the fixation, and its distance to the fixation was shorter than that of the comparator dots. In contrast, when the shifting distance was positive, the center dot was farther away from the fixation, and its distance to the fixation was longer than that of the comparator dots. This set of three dots appeared in one of eight locations in the peripheral visual field, as shown in Figure 4.3.

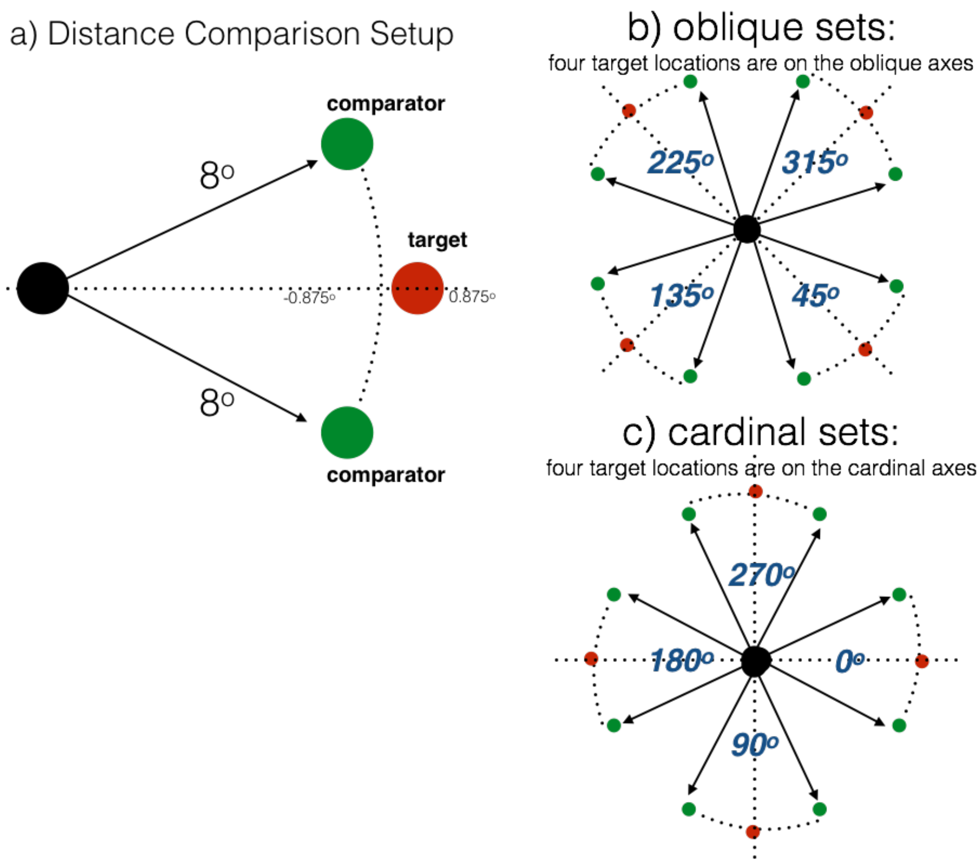


Figure 4.3 Stimulus configurations in Experiment 4.2

Procedure

Participants were asked to report whether the distance between the fixation and target dots was longer or shorter than the distance between the fixation and flanking dots. Stimuli were generated and presented using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Each trial started with a fixation of 500 ms. The display of three dots was randomly presented at one of the eight possible locations on the screen, lasting for 80 ms. The trial ended when participants responded to the display. Participants were instructed to respond as accurately as possible.

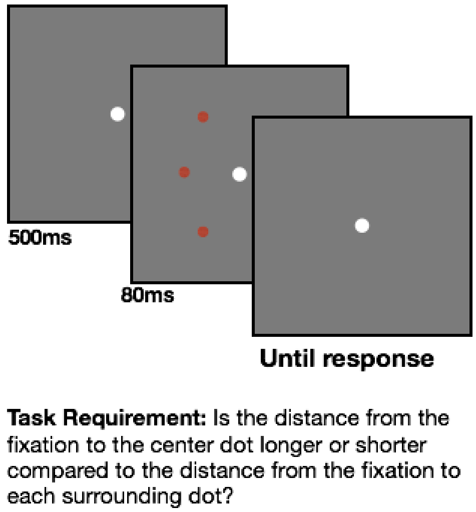


Figure 4.4 Experimental procedure in Experiment 4.2

Results

We excluded data from four participants from further analysis because their performance appeared random and showed a strong response bias in choosing one option over the other. Data from the remaining 18 participants were included in further analysis.

We fitted a psychometric function in each Vernier discrimination task. We used the python package *psignifit* 4.0 (Schütt et al, 2016) and chose the cumulative normal function as the model of psychometric function. *Psignifit* 4.0 provides a full Bayesian estimation of parameters in psychometric functions. We used the maximum a posterior (MAP) estimation as the point of subjective equality (PSE) of distance: the position of the target dot in the Vernier task at which people perceived the target dot as having a greater distance than the comparator dots in 50% of trials. We also extracted the 95% confidence interval of PSE based on its posterior distribution.

Psychometric Functions of Pooled Responses

The psychometric functions of pooled responses showed different PSEs when the dot set was placed on oblique axes and on cardinal axes, as illustrated in Figure 4.5. When the dot set was placed on an oblique axis, such as 45º, 135º, 225º, or 315º, in the peripheral visual field, people's distance judgment was accurate. The target point shift distance at PSE was near 0, and the 95% confidence interval of the PSE was included with 0 shift distance. When the dot set was presented on a cardinal axis, such as on the horizontal or vertical meridian, there was a perceptual bias: people perceived a shorter physical distance of the target to be the same as the distances of comparator dots, as if the target was dragged towards the straight line that connected the two comparator dots. This distance judgment distortion only appeared when the set was presented on cardinal axes and comparator dots formed a horizontal or vertical line.

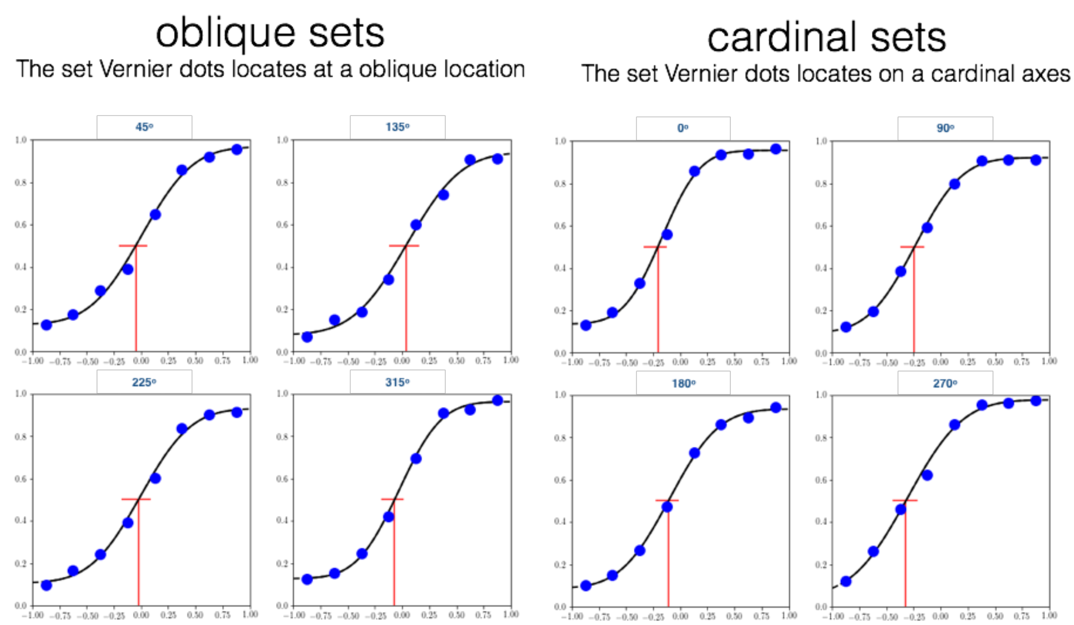


Figure 4.5 Psychometric Functions of Equal Distance Perception at Different Target Direction

Vernier Set Location: oblique	Vernier Set Location: cardinal

(45°, 135°, 225°, 315°)			(0°, 90°, 180°, 270°)		
location angle	Perceptual Bias Magnitude	95% Confidence Interval	**location angle**	Perceptual Bias Magnitude	95% Confidence Interval
45°	-0.049 ° (vis. deg.)	[-0.243, 0.086]	0°	-0.207° (vis. deg.)	[-0.373, -0.102]
135°	0.035° (vis. deg.)	[-0.152, 0.173]	90°	-0.242° (vis. deg.)	[-0.395, -0.131]
225°	-0.024° (vis. deg.)	[-0.220, 0.106]	180°	-0.108° (vis. deg.)	[-0.259, 0.000]
315°	-0.069° (vis. deg.)	[-0.221, 0.034]	270°	-0.321° (vis. deg.)	[-0.477, -0.196]

Table 4.1 Psychometric function fitting results of pooled response data in Experiment 4.2

Individual Fittings and Hypothesis Testing

To confirm our previous observation with statistical testing, we fitted a psychometric function for dot-set location for each participant. We extracted the target point shift distance at each PSE point. Our null hypothesis was that perceptually equal distances were consistent with physically equal distances. This null hypothesis predicted that the middle point shift distance at each PSE point was not significantly different from 0. In contrast, our alternative hypothesis was that perceptually equal distances were not consistent with physically equal distances. We pooled PSE from four cardinal-location conditions together and from four oblique-location conditions together, and performed three *t-tests* to examine 1) whether the PSE in cardinal-location conditions was significantly biased away from the physical truth (independent t-test), 2) whether the PSE in oblique-location conditions was significantly biased away from physical truth (independent t-test), and 3) whether the PSE in two conditions were significantly different from each other (pairwise t-test). We corrected for multiple comparisons by setting $\alpha = 0.0167$ ($\alpha = 0.05 / 3$ comparisons). We also used a stricter critical value $\alpha = 0.003$ ($\alpha = 0.01 / 3$ comparisons) to reject the null hypothesis.

The results showed significant misjudgment of equal distances when a dot set was presented on the cardinal axes: when the target point and the comparators had perceptually equal distance to the fixation, the target dots actually had a shorter physical distance than the comparators did ($t(17) = -6.196$, $p = 3.35 \cdot 10^{-8}$). The PSE from the oblique conditions was not significantly different from 0, suggesting that when a dot set was presented on the oblique axes, people did not significantly misjudge the distance ($t(17) = -0.538$, $p = 0.592$). In addition, there was a significant difference between the PSEs in the cardinal conditions and those in the oblique conditions ($t(17) = -10.656$, $p = 2.23 \cdot 10^{-16}$), suggesting that people might have used different perceptual computations when they compared the target's distance with the comparators' distances in different conditions. Figure 4.6 provides a schematic description of our results.

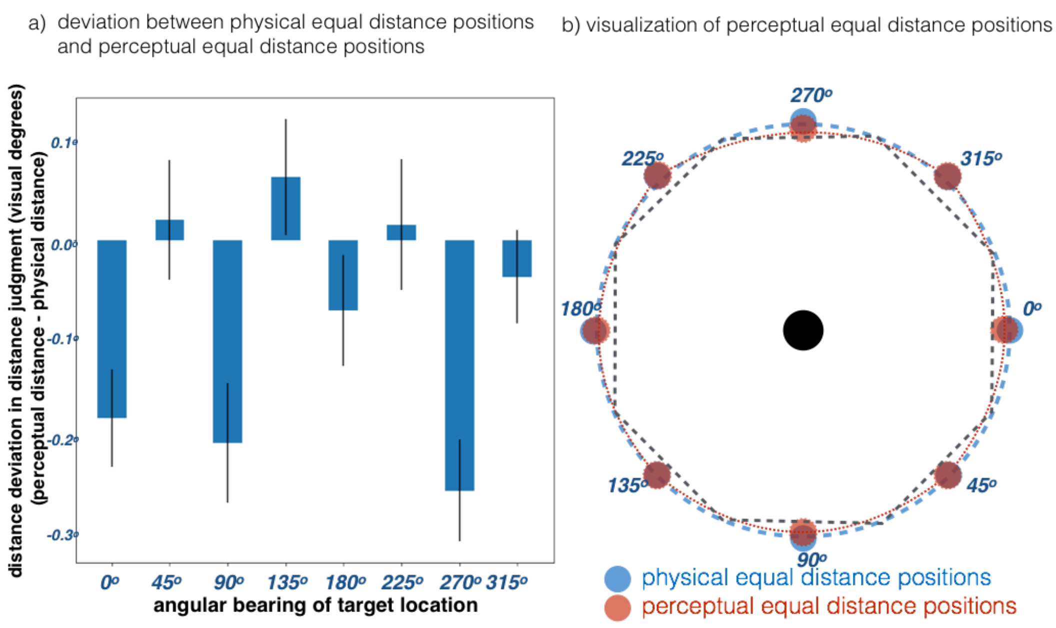


Figure 4.6 PSE of equal distance points

Discussion

In this experiment, we tested whether distance is a primitive in the visual representation of positions. We argued that if this is the case, distance judgment should be accurate. Our results showed that when people compared distance in oblique locations, they were accurate; but when they compared distance in cardinal locations, that is, on the horizontal and vertical meridians, their perception was systematically biased. The accurate distance judgment in oblique locations suggests that distance can be directly read out and used in comparison regardless of angular bearings. For example, in one comparison, the target dot had an angular bearing of 45°, whereas the two comparators had angular bearings of 32.5° and 67.5°, respectively. Even though these three positions appeared at different angles, participants accurately perceived and compared their distances. Thus, distance information should be a primitive in position representation in the comparison computations, and it is represented independently of angle information.

However, when the target was presented on the cardinal axes, participants failed to accurately compare distances between the target and the comparators. In these cases, a shorter distance of the target on the axis was considered to be equal to a longer distance of comparators off the axis. There are two possible reasons for this perceptual bias. First, participants might have failed to perceive the distance accurately when the target was placed on the axis. Second, participants may have perceived the distance accurately, but the perceptual bias might have been caused by the comparison computation between the target and the comparator. When the target was placed on a cardinal axis, the comparator-target set formed a curve that was oriented either vertically or horizontally. After perceiving this dot set, participants might have used heuristics in distance comparison. For example, when the target was placed on the positive side of the horizontal axis (at 0° angle condition), the comparator-target set formed a vertically oriented curve. Instead of reading out the exact distance information of these three dots, participants might have directly read out the horizontal displacement of each dot, and heuristically compared differences in horizontal displacements to determine whether or not the target had a longer distance than the comparators. People are very good at perceiving vertical and horizontal orientations (Appelle, 1972). When they can reduce the two-dimensional distance comparison to a one-dimensional displacement comparison without large errors, they may choose to read out the one-dimensional displacement information and use the solution of this simple comparison task to answer the original distance comparison question. Thus, as illustrated in Figure 4.6, when the target was on the cardinal axes, the subjective equal distance position of the target was biased towards the position that had the same one-dimensional displacements as the comparators. When the comparator-target set was placed on oblique axes, the set formed a curve that was oriented obliquely. In these cases, the distance comparison task could not heuristically be reduced to a one-dimensional comparison task. Thus, participants needed to read out exact distance information to compare the distance. As a result, participants showed accurate judgments.

Experiment 4.3a Aligning Positions on a Straight Line

In the third experiment, we tested whether translation information is primitive in the visual position representation. We asked participants to align three dots on a straight line as accurately as possible. If a Cartesian coordinate system is used and translation is primitive, we would expect participants to accurately align dots on straight lines regardless of cardinal or oblique orientation. Alternatively, if a polar coordinate system is used, we might expect participants to show perceptual biases in alignment performance. If they could use heuristics in the alignment task, as suggested in Experiment 4.2, they might reduce the alignment task to a one-dimensional displacement comparison task when stimuli were oriented vertically or horizontally, and thus show less perceptual bias.

Method

Participants

A group of 21 Johns Hopkins University undergraduate and graduate students participated in exchange for course credit. All participants reported normal or corrected-to-normal visual acuity and gave written informed consent. The experiment was approved by the Johns Hopkins University institutional review board (IRB).

Stimulus

A set of three dots, each of which measured 0.5° in diameter, formed a curve on the screen, as shown in Figure 4.3. The two flanking dots were placed at 8.94° away from the fixation and 8° away from each other. Thus, the middle of the two flanking dots appeared 8° away from the fixation. The center dot was placed in between, but shifted along a line that was perpendicular to the line that connected the two flanking dots. The shifting distance was from -0.45° to 0.45°, with a step size of 0.1°. The corresponding distance between the center dot and the fixation point varied from 7.55° to 8.45°. That is, when the shifting distance was negative, the center dot was closer to the fixation than the middle between the two flanking dots; and when the shifting distance was positive, the center dot was farther away from the fixation than the middle between the two flanking dots. In the former case, the three dots formed a curve that pointed towards the fixation; in the latter case, the three dots formed a curve that pointed away from the fixation. Similar to Experiment 4.2, this set of three dots appeared in one of eight locations in the peripheral visual field.

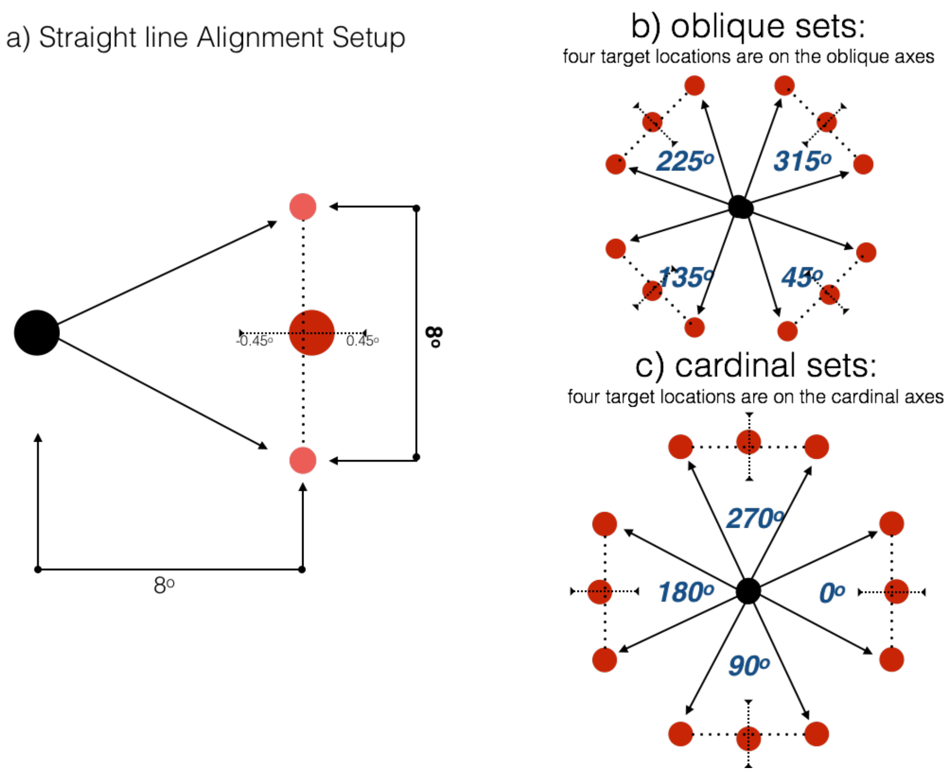


Figure 4.7 Stimulus configurations in Experiment 4.3a

Procedure

Participants were asked to report whether the curve consisting of the three dots pointed towards or away from the fixation. Stimuli were generated and presented using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Each trial started with a fixation for 500ms. The display of three dots was then randomly presented in one of the eight possible locations on the screen, lasting for 80ms. The trial ended when participants responded to the display. Participants were instructed to respond as accurately as possible.

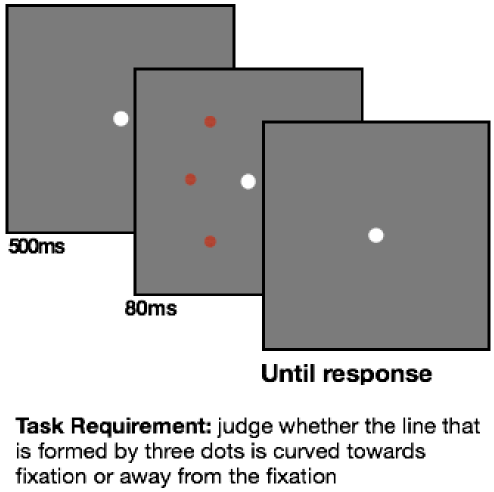


Figure 4.8 Experiment procedure in Experiment 4.3a

Results and Discussion

We excluded data from two participants from further analysis because their performance appeared random and could not be fitted reasonably by psychometric functions. Data from the remaining 19 participants were included in further analysis.

Using a cumulative normal function as the model, we fitted psychometric functions to people's responses. We used the maximum a posterior (MAP) estimation as the point of subjective equality (PSE) of a straight line: the position of the middle dot in the Vernier stimuli at which people perceived that the curve made by the Vernier dot set faced away from the fixation in 50% of the trials, and perceived that this curve faced towards the fixation in the other 50% of the trials. We also extracted the 95% confidence interval of the PSE based on its posterior distribution.

First, we pooled response data from all participants together and fitted a single psychometric function to all responses in the Vernier discrimination task at each of the eight locations. We also separately fitted psychometric functions to responses in each Vernier discrimination task for each participant, and used the MAP to extract a PSE of straight line perception in each psychometric function. We then performed hypothesis testing on these data to investigate whether a perceptual straight line was the same as a physical straight line, or whether there was a systematic perceptual bias in straight line perception.

Psychometric Functions of Pooled Responses

Psychometric functions of pooled responses showed that there were systematic biases in perceiving a straight line. The middle point shift distance at PSE was larger than 0 in all cases, and the 95% confidence interval of the PSE in 6 of 8 conditions did not overlap with 0 shift distance. These results showed that people perceived a curve that faced away from fixation to be a straight line. This distorted straight line perception was biased towards the circumference of a fixation-centered circle that crossed the two flanking points in a Vernier dot set.

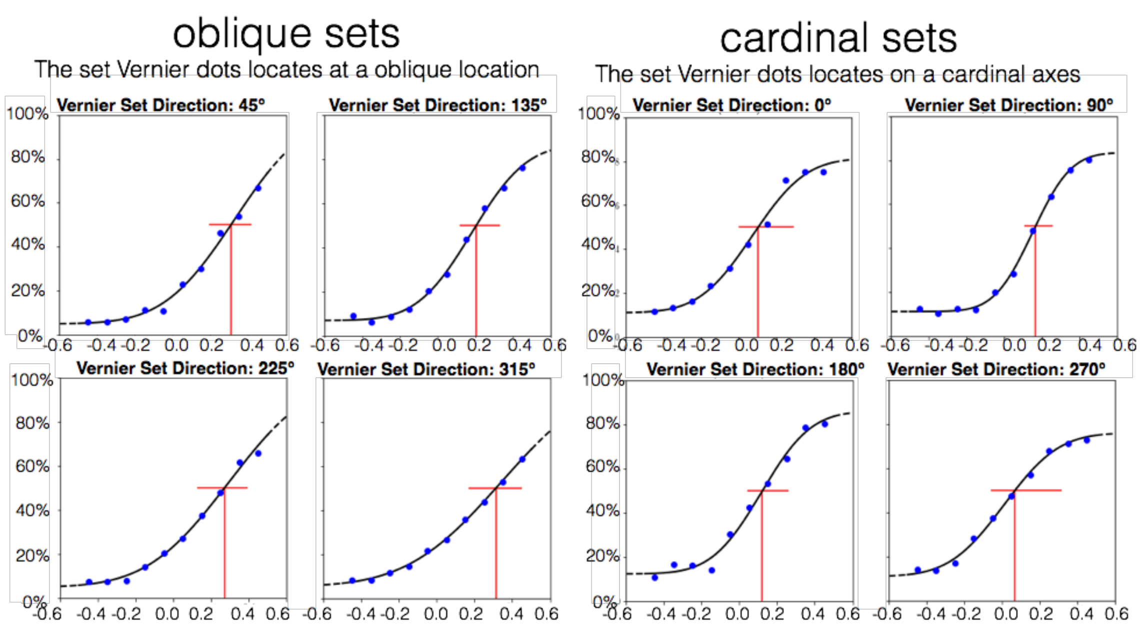


Figure 4.9 Psychometric functions of straight line perception

Direction	Middle point shift distance at PSE	95% confidence interval of the PSE	Direction	Middle point shift distance at PSE	95% confidence interval of the PSE
0°	0.103° (vis. deg.)	[-0.011, 0.327]	45°	0.309° (vis. deg.)	[0.174, 0.435]

90°	0.168° (vis. deg.)	[0.101, 0.277]	135°	0.204° (vis. deg.)	[0.106, 0.358]
180°	0.121° (vis. deg.)	[0.028, 0.297]	225°	0.272° (vis. deg.)	[0.105, 0.419]
270°	0.069° (vis. deg.)	[-0.076, 0.359]	315°	0.313° (vis. deg.)	[0.139, 0.481]

Table 4.2 Psychometric function fitting results of pooled response data in Experiment 4.3a

The results also showed that perceptual biases were smaller when the Vernier set was present on the cardinal axes than when it was present on the oblique axes. When the Vernier set was present at an angular direction of 0° or 270°, the 95% confidence interval of the PSE estimation included 0-shift magnitude. When the Vernier set was presented at an angular direction of 90° or 180°, the estimated PSE bias was also smaller in magnitude than bias in other oblique conditions.

Individual Fittings and Hypothesis Testing

To confirm the above observation with statistics, we fitted a psychometric function for each Vernier discrimination task for each participant, and extracted the middle point shift distance at each PSE point. Our null hypothesis was that perceptual straight lines correspond to physical straight lines if a Cartesian coordinate system is used in visual spatial representation. This null hypothesis predicted that the middle point shift distance at each PSE point would not be significantly different from 0. Conversely, our alternative hypothesis was that a perceptual straight line does not correspond to a physical straight line due to perceptual biases that stem from a non-Cartesian coordinate system used in visual spatial representation. Thus, for each one of eight different Vernier discrimination tests, we conducted eight one-sample *t-tests*, and corrected for multiple comparisons by setting $\alpha = 0.00625$ ($\alpha = 0.05$ / 8 comparisons). We also used a stricter critical value $\alpha = 0.00125$ ($\alpha = 0.01$ / 8 comparisons) to reject the null hypothesis. The results showed significant perceptual biases in straight line perception when a Vernier set was presented in all four oblique conditions (**45°** ($t(18) = 6.94$, $p = 1.76 \cdot 10^{-6}$), **135°**($t(18) = 4.71$, $p = 1.71 \cdot 10^{-4}$), **225**°($t(18) = 5.85$, $p = 1.52 \cdot 10^{-5}$, and **315°**($t(18) = 7.02$, $p = 1.49 \cdot 10^{-}$

[6])), but only in one cardinal condition (**90°**(t(18) = 5.65, p = 2.29 • 10^{-5})). The results also indicated a marginally significant perceptual bias when the Vernier set was presented in the other three cardinal angular directions (**0°** (t(18) = 3.18, p = 0.0052), **180°**(t(18) = 3.89, p = 0.0011) **270°** (t(18) = 2.94, p = 0.0087)). An additional paired t-test confirmed a significant difference in biases between the cardinal and oblique conditions (t(18) = -4.19, p = 7.40 • 10^{-5}).

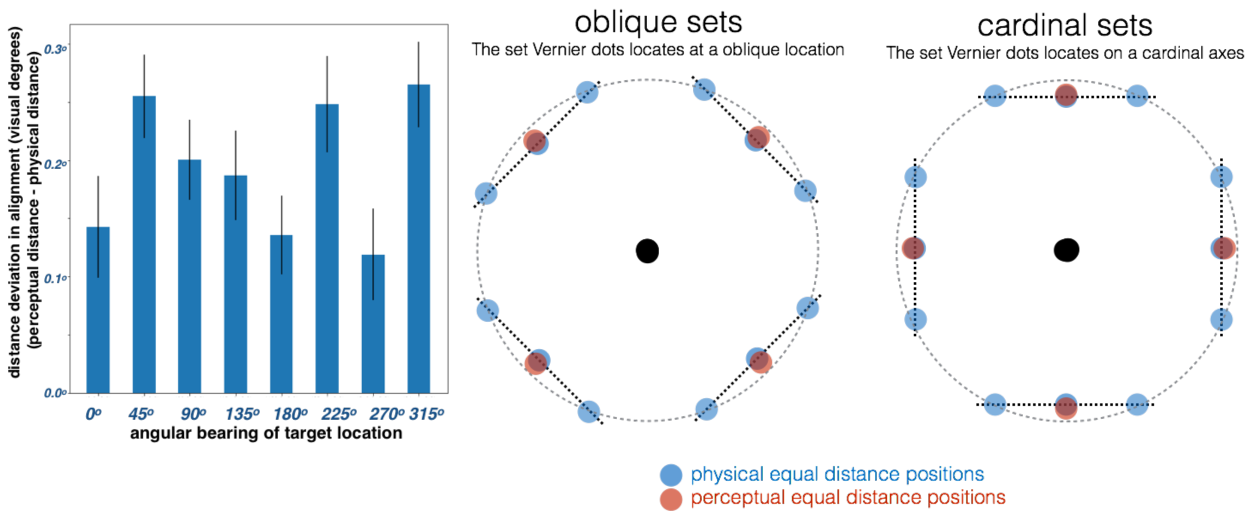


Figure 4.10 PSE of straight lines in each Vernier discrimination condition

Surprisingly, a perceptual straight line was not physically straight. People had a systematic bias: for participants to perceive a straight line, we had to move the middle dot of a Vernier set farther away from the fixation than a physical straight line required. This result indicates that translational information is not a primitive in position representations. When the Vernier set was presented obliquely, the perceptual bias was significant, suggesting that translational information was not computed accurately. Recall that when a similar dot set was present obliquely, participants could accurately compare the distances of each dot in an equal distance comparison task in Experiment 4.2. Comparing these two cases, perceptual inaccuracy in the present experiment may have come from a polar coordinate system that uses distance and angle information to compute the straight line alignment. To describe a straight line with distance and angle information requires fitting a complex, non-linear function. The visual system may approximate this non-linear function with simple algorithms. For instance, in this experiment, the visual system may have heuristically conducted an equal distance comparison, thus the perceptual straight line was biased towards the circumference of a circle on which positions had equal distances to the fixation. Even when the Vernier set was presented vertically or horizontally, the perceptual alignment was still biased away from the physical straight line. Participants did not completely reduce the alignment task to a one-dimension displacement comparison task, nor could they directly access information about horizontal or vertical displacement. In sum, the results of this experiment suggest that primitives from a Cartesian coordinate system are not used in position alignment computations. Instead, a polar coordinate system may be used and be responsible for perceptual bias in the alignment task.

Experiment 4.3b Aligning Positions on Straight Lines of Different Lengths

If a polar coordinate system is used in visual perception, and if the perceptual bias in Experiment 4.3b comes from an algorithm that approximates Cartesian straight lines with primitives in a polar coordinate system, we would expect biases to increase as the distance between the two flanking dots increases. When this distance increases, the two flanking dots simultaneously and symmetrically become farther away from the fixation. As a result, these two dots appear on a larger circle that is centered on the fixation. This larger circle may bias perception of the middle dot even farther away from the true positions when people try to perceive a straight line, as illustrated in Figure 4.7. In contrast, if this bias is due to a noisy Cartesian coordinate representation, we would not expect a perceptual bias increase in our manipulation, because the stimuli dots are still on the same straight line.

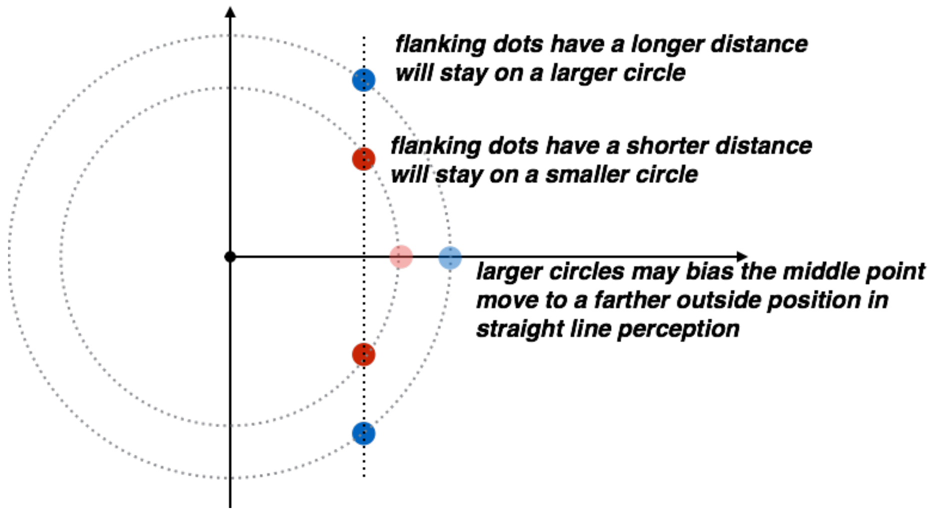


Figure 4.11 A polar coordinate system predicts that larger distances between the two flanking dots can cause larger perceptual biases in the straight line perception

We test the above hypothesis in the following. In a new experiment, we investigated whether the bias increases depending on the distance between the two surrounding dots. The logic is that the larger the distance between the two surrounding dots is, the larger the circumference of the circle that is determined by the fixation and the surrounding dots will be, which in turn results in a larger bias towards the circumference.

Method

Participants

Two groups of Johns Hopkins University undergraduate students participated in the experiment in exchange for course credits. Each group consisted of 21 participants. All participants reported normal or corrected-to-normal visual acuity and gave written informed consent. The experiment was approved by the Johns Hopkins University institutional review board (IRB).

Stimulus

Stimuli in this experiment were still a set of three dots, each of which measured 0.5° in diameter. As illustrated in Figure 4.12, the distance between the two flanking dots varied; it could be 4°or

8°. The middle of the two flanking dots was always 8° away from the fixation. Thus, the distance between the flanking dots and the fixation could be 8.24° or 8.94°. Accordingly, the circumference that was determined by the fixation and flanking dots increased as the distance between the two flanking dots increased. The center dot appeared in between, but shifted along a line that was perpendicular to the line that connected the two flanking dots. The shifting distance was also from -0.45° to 0.45°, with a step size of 0.1°.

The three-dot Vernier set appeared in different locations in two conditions, as illustrated in 4.12. In an oblique condition, the set appeared in one of four oblique locations (4.12b); while in a cardinal condition, the set appeared in one of four cardinal locations (4.12c).

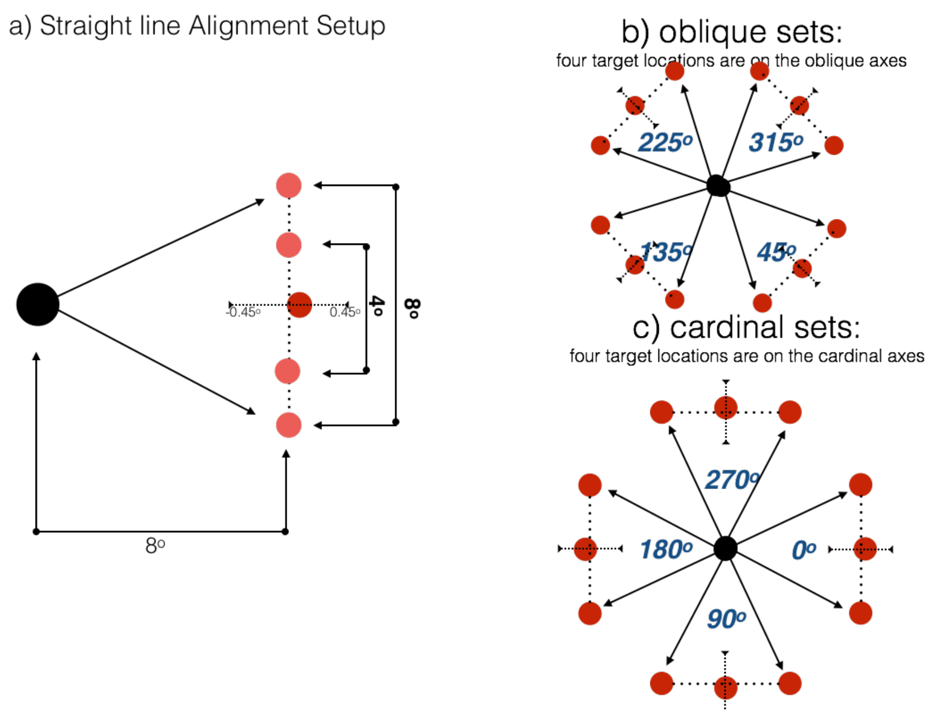


Figure 4.12 Stimulus configurations in Experiment 4.2b

Procedure

Participants were assigned either to the oblique condition or to the cardinal condition at the beginning of the experiment. Subsequent procedures in this experiment were the same as those in

Experiment 4.3a: participants perceived a set of three dots that flashed on the screen for 80ms and reported whether the curve made of the three dots pointed towards or away from the fixation.

Results

We excluded data from 4 participants in the cardinal condition and 3 participants in the oblique condition from further analysis because their performance appeared random and could not be fitted reasonably by psychometric functions. Data from the remaining 17 participants in the cardinal condition and 18 participants in the oblique condition were included.

In the analysis, we used a cumulative normal function as the model of psychometric function to fit discrimination responses in each Vernier discrimination task, and extracted PSE and its 95% confidence interval. We also started with a descriptive analysis by pooling response data from all participants together and fitting psychometric functions to these pooled data. Then, we fitted psychometric functions for individual performances and performed inference hypothesis testing. To understand when and what coordinate system was used in the alignment computation, we conducted mixed measures ANOVA of the effects of flanking distance (4° or 8°) and dot-set locations (cardinal vs. oblique) on discrimination thresholds.

Figure 4.13 illustrates the psychometric functions fitted to pooled responses from all participants from the cardinal and oblique conditions. In both conditions, the perceptual bias was small when the flanking distance was 4°. When the flanking distance was 8°, the perceptual bias was larger, similar to that in Experiment 4.2a. When the Vernier set was placed at locations in the cardinal condition, the magnitude of the perceptual bias was similar whether the flanking distance was 4° or 8°. When the set was placed at locations in the oblique condition, however, the magnitude of perceptual bias increased clearly when the flanking distance changed from 4° to 8°.

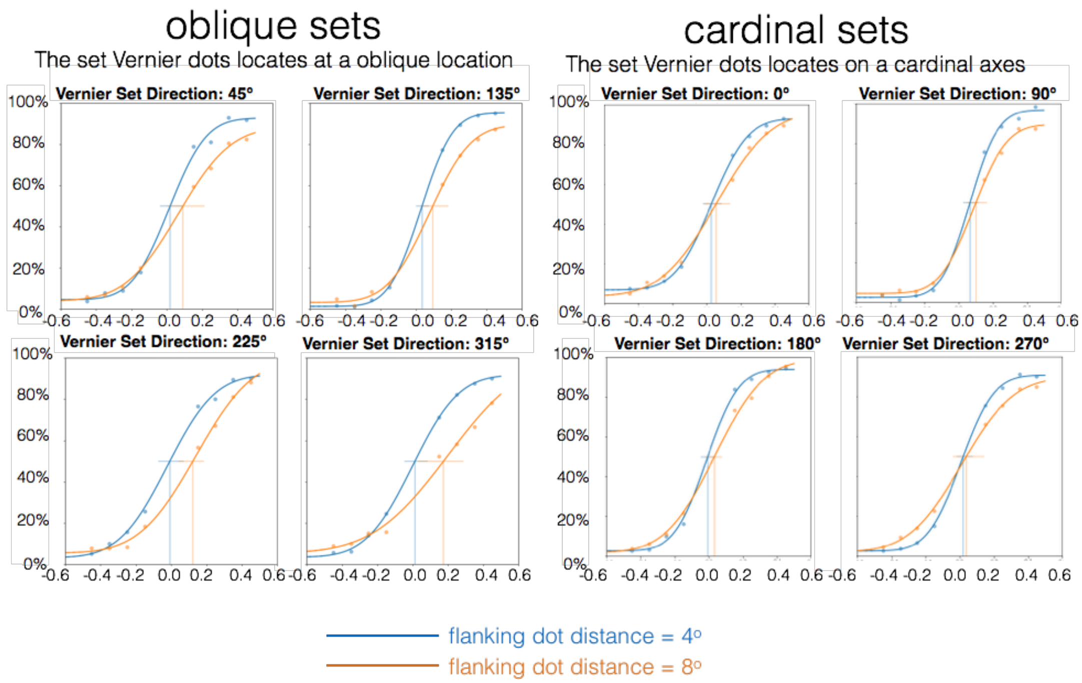


Figure 4.13 Psychometric functions of straight line perception

To further confirm the fitting results in the pooled psychometric function, we fitted a psychometric function in each alignment task for each participant and conducted a 2 x 2 mixed design ANOVA (flanking distance: 4° vs. 8°, within-subject; location condition: cardinal vs. oblique, between-subject) on the magnitude of perceptual bias at the PSE. As illustrated in Figure 4.10, the results showed a significant interaction effect between flanking distances and location conditions ($F(1, 33) = 16.16$, $p = 3.18 \cdot 10^{-4}$), suggesting that the magnitude of perceptual bias when people aligned a cardinally oriented straight line was less affected by the flanking distance than an obliquely oriented straight line. The results also showed a significant effect of distance ($F(1, 33) = 39.25$, $p = 4.43 \cdot 10^{-7}$), indicating that the larger the flanking distance was in the Vernier set, the larger the perceptual bias would be. On the other hand, the effect of the location condition was not significant ($F(1, 33) = 1.46$, $p = 0.235$), suggesting that participants showed similar perceptual bias in both the cardinal alignment task and the oblique alignment task.

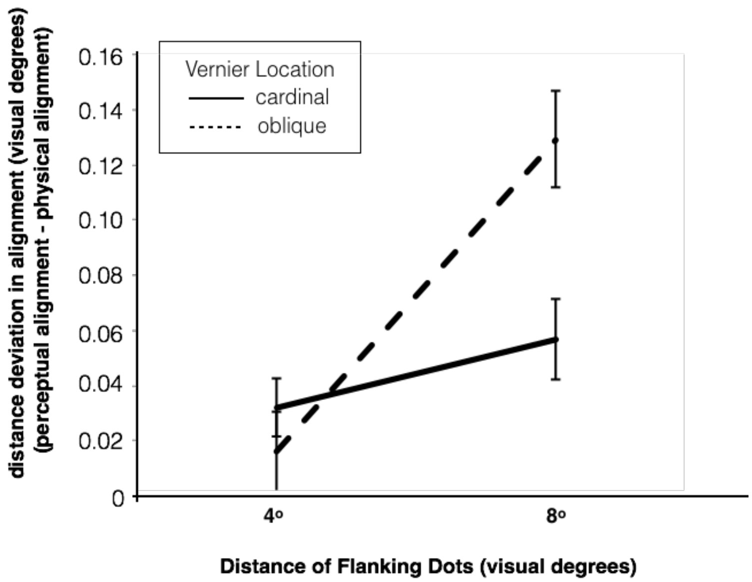


Figure 4.14 Effects of distance of flanking dots on the magnitude of perceptual bias in straight line alignment task

Discussion

Our results show that participants had similar perceptual biases in both oblique and cardinal alignment tasks. Consistent with Experiment 4.3a, the present results suggest that translation information in a Cartesian coordinate system is unlikely to be used in the visual perception. Our results further indicate that the magnitude of this bias is larger when the flanking distance of the Vernier set is larger. These results are consistent with the polar hypothesis, as illustrated in Figure 4.7: the larger the distance between the two flanking dots is, the larger the circle where these two dots are located will be. A larger circle leads to a larger straight line misperception by biasing the middle point closer to the circumference of the circle, but farther away from the fixation. The significant interaction effect indicates that perceptual bias was affected more by the flanking distance of the Vernier set when the set was placed obliquely. This could be because participants could only use distance and angle information in the polar representation to perform the alignment task. Thus, their performance was strongly affected by the biases in the polar computation in Figure 4.11. In contrast, when the set was placed vertically or horizontally, the perceptual bias was less affected by the flanking distance. Participants may have used heuristics to align straight lines vertically or horizontally, similar to the heuristic hypothesis in previous experiments. They may have reduced the straight line alignment task to a one-dimensional displacement comparison task in these scenarios and used polar primitives to calculate this one-dimensional information with less computational bias. When the flanking distance increased, calculation of the one-dimensional displacement magnitude might not have been as strongly affected as the calculation of two-dimensional alignments.

Chapter 5 Formats of Representation Affect Complexity in Perceptual Computation

In his discussion of representations, Marr (1982) pointed out that "how information is represented can greatly affect how easy it is to do different things with it." In this chapter, we rely on the logic that the ease of executing specific computations depends on representational format, both in general and in the case of positions in pattern perception. To perceive a geometric feature such as a straight line, a Cartesian coordinate system requires fitting a simple linear function, $y = ax + b$, while a polar coordinate system requires fitting a complex, non-linear description, $\rho = \frac{b}{\sin\theta - a\cos\theta}$. In contrast, perceiving a circle requires a simple, linear description in a polar coordinate system, such as $\rho = 3$, but a non-linear description in a Cartesian coordinate system, such as $x^2 + y^2 = 3$.

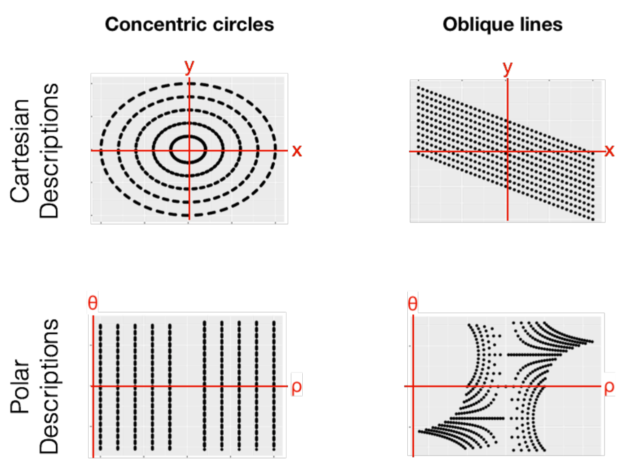


Figure 5.1 Geometric patterns have different descriptions in different coordinate systems

Fitting a linear model is less computationally complex than fitting a non-linear model, especially when noise is introduced in the computational process. Machine learning research has shown that simple models are easier to fit and behave more robustly than complex models when data is noisy. A fitting algorithm also converges quickly to optimal results. Complex models, in contrast, are more difficult to learn and tend to over-fit noise in data. They often result in large variance when different noise is embedded in data. A complex algorithm also tends to converge more slowly and it can become stuck at some local optimal solutions without finding global optimal results (Friedman, Hastie, & Tibshirani, 2001).

Similarly, although the concept of psychological complexity can appear vague, it has been characterized mathematically in some domains, and it has even been shown that more complex functions are more difficult to learn. For example, in the case of a simple Boolean concept learning task, people's reaction times increase as the complexity of a Boolean concept increases (Feldman, 2003). In this chapter, we apply a similar logic: if geometric patterns with simple Cartesian descriptions are easier to perceive in the presence of noise, then the visual system probably uses a Cartesian coordinate system format; likewise, if patterns with simple polar descriptions are easier to perceive, the visual system likely uses a polar format.

To test our hypothesis, we designed experiments in which people detected static and motion geometric patterns constructed by random dots. For static patterns, we adapted Glass patterns. A Glass pattern is made by superimposing a set of randomly distributed features (e.g. small dots) on a geometrically transformed copy of itself, illustrated in Figure 5.2 (Glass 1969, 2002; Prazdny, 1984). Superimposing different transformations leads to the appearance of different patterns, such as concentric circles, translational lines in parallel, radial lines expanding from an origin, or hyperbolic curves (Glass & Switkes 1976). For motion patterns, we adapted random dot coherent motion stimuli. These stimuli are defined by spatiotemporal correspondence among a set of dots randomly distributed in space. When dots move in the same direction or the same manner, the stimulus gives rise to the appearance of a global geometric pattern.

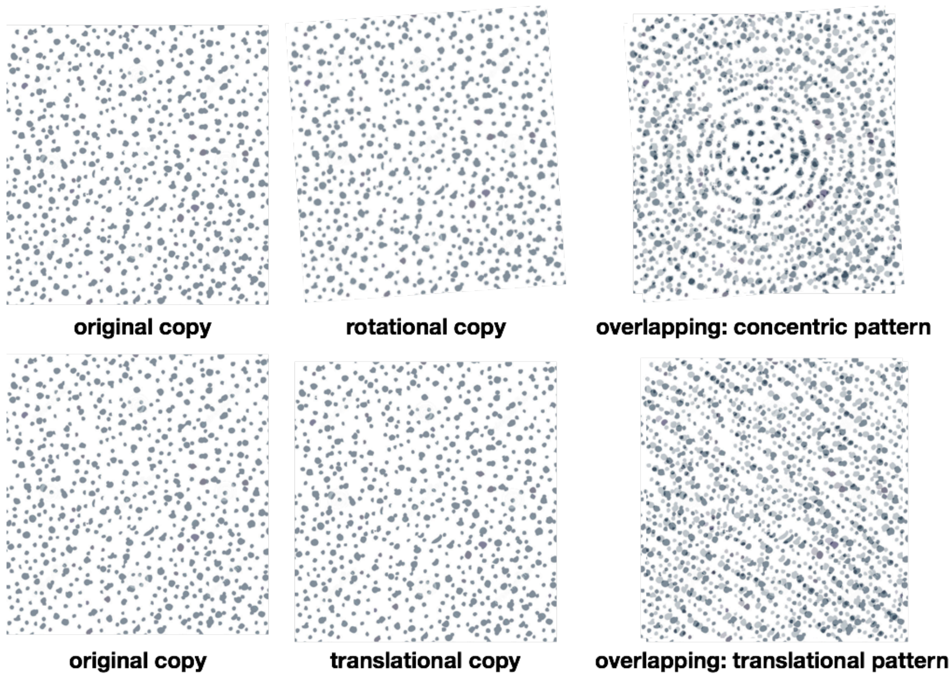


Figure 5.2 Glass pattern examples: concentric and translational patterns

Random dot stimuli were especially suited to this study for two reasons. First, the only information that our visual system can use to discriminate a geometric pattern is spatial correspondences between dots. These spatial correspondences rely heavily on how positions are represented. No explicit Gestalt grouping cue or local object information helps to construct a global pattern. Local features of these stimuli, such as dot density and dot distance, are also matched and controlled among all stimuli. Second, we could have precise control of signal-to-noise ratio for each stimulus. Varying the proportion of dots that have the same spatial correspondence regularity and perturbing the rest randomly in space, we could systematically vary the strength of signals that defined geometric patterns, thus introducing different levels of perceptual uncertainty. Using these random dot stimuli, we tested (1) how well people detect various geometric patterns in noise and (2) what patterns are easier to detect than others when two patterns exist at the same time.

Experiment 5.1 Detecting Patterns Among Noise

In this experiment, we examined the perception of geometric patterns by asking people to perceive patterns embedded in noise. We added different proportions of noise into each pattern and measured the minimal amount of signal required to reliably detect each pattern against pure noise.

Methods

Participants

Two groups of Johns Hopkins University undergraduate and graduate students participated in the experiment in exchange for course credit. Each group consisted of 21 participants. All participants reported normal or corrected-to-normal visual acuity and gave written informed consent. The experiment was approved by the Johns Hopkins University institutional review board (IRB).

Apparatus

All experiments took place in a dim, sound-attenuated room. There was no light source except for a computer monitor. All stimuli were presented on a Macintosh iMac computer with a refresh rate of 60 Hz. The viewing distance was 60 cm, so that the display subtended 39.43° × 24.76° of visual angle.

Stimulus

1. Static Patterns

Stimuli were variations of Glass patterns (Glass, 1969), as shown in Figure 5.3. In a Glass pattern stimulus, a randomly generated set of 1,000 dots was overlaid on top of a geometrically transformed copy of itself, forming a global pattern of 2,000 dots in space. The geometric transformation could be a translation, an expansion, or a rotation of the original set. Each dot measured 0.02° by 0.02° (1 *pixel* by 1 *pixel*), in a figure. The figure measured 6.9°x 6.9° visual angles in size.

Five types of Glass patterns were used as stimuli, as illustrated in Figure 5.3. 1) The ***concentric*** pattern was made by overlapping a set of random dots and its rotational copy, resulting in the appearance of multiple concentric circles with different radii. 2) The ***radial*** pattern was made by overlapping a set of random dots and its expansion, resulting in the appearance of multiple radial lines that all passed through the image center. 3) The ***translational*** pattern was made by overlapping a set of random dots with its translational copy, resulting in multiple parallel 45°-oriented lines. 4) The ***concentric-part*** pattern, which resembled the **concentric** pattern except that the center of the circle was outside of the figure boundary, appeared to be a cropped part of an intact concentric pattern. 5) The ***radial-part*** pattern, which resembled the radial pattern except that the intersection point of all the radial lines was outside of the image boundary, appeared to be a cropped part of an intact radial pattern.

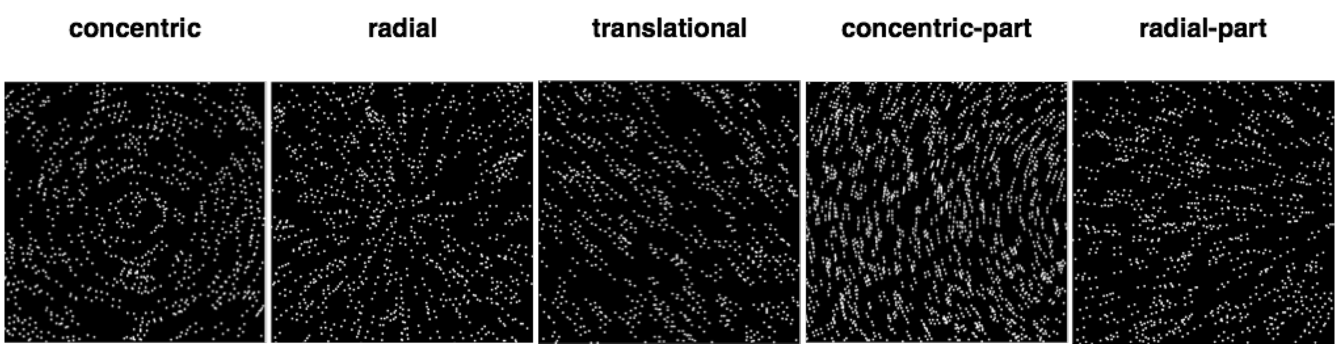


Figure 5.3 Five Glass patterns in Experiment 5.1

2. Motion Patterns

We also produced five moving patterns. Each pattern consisted of 1,000 random dots, with each dot measuring 0.02° by 0.02° (1 *pixel* by 1 *pixel*). These dots moved rotationally, radially, or translationally in a square aperture of 6.9°x 6.9° visual angles in size, resulting in motion versions of the **concentric** pattern, **radial** pattern, **translational** pattern, **concentric-part** pattern, and **radial-part** patterns.

3. Signal Strength

To introduce perceptual uncertainty, we manipulated the proportion of signals and noise in each figure, varying signal strength from 5% to 75%, with a step size of 10%.

In each static pattern, the signal strength was defined as the proportion of dot correspondence that defined the pattern, and the remaining dots were randomly perturbed to form perceptual noise. For instance, if we set 25% dot correspondences to be randomly oriented, the signal strength became 75% because only three quarters of the dot correspondences obeyed the transformation that defined the pattern.

In each motion pattern, the signal strength was determined by the proportion of dots that moved coherently to give rise to the pattern perception. For example, a stimulus with 35% of its dots moving translationally and the other 65% of dots moving randomly was defined as a translational motion pattern of 35% signal.

Design and Procedure

Stimuli were presented at the center of the screen. Each pattern was presented 12 times at each signal strength level, resulting in 480 trials in total (5 signal patterns × 8 signal strength × 12 repetitions). Each stimulus was randomly generated at the beginning of a trial. Therefore, even if two stimuli formed the same type of pattern, such as two concentric Glass patterns, their local features and random dot locations were different. Stimuli were generated and presented using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

At the beginning of each block, a figure of a target pattern in this block was presented with 100% signal strength. Participants were allowed to view the figure until they understood the pattern, but they were required to maintain fixation on the center dot of the screen. Each trial began with a 500ms fixation point. Subsequently, two figures were successively presented on the screen, each of which lasted for 150ms with an inter-stimulus interval of 500ms. One of the figures contained the target pattern, whereas the other only contained randomly distributed noise. Participants were required to report which of the two figures contained the target pattern. This was a two-alternative forced choice (2AFC) discrimination task; a random guess performance would result in 50% discrimination accuracy. A trial ended when the response was recorded.

Analysis

We fitted a psychometric function to discrimination accuracies across all signal strength levels for each pattern in the experiment. We used the python package *psignifit* 4.0 (Schütt et al, 2016) and chose the Weibull function as the model of the psychometric function. *Psignifit* 4.0 provided a full Bayesian estimation of parameters in psychometric functions, and we used the maximum a posterior (MAP) estimation as the point estimation of the ***discrimination threshold***, a signal strength level at which people achieved 75% discrimination accuracy, based on fitting. We also extracted the 95% confidence interval of the discrimination threshold estimation based on its posterior distribution from the full Bayesian estimation process.

First, we pooled response data from all participants together and fitted a single psychometric function to all responses to each pattern. This pooled fitting process resulted in five psychometric functions that described the performance of discriminating each noisy pattern from a random-noise figure. Second, we separately fitted psychometric functions to responses to each geometric pattern for each participant and used MAP to extract a point estimation of the discrimination threshold in each psychometric function. We then used repeated measures ANOVA to test whether there were differences in perceiving different geometric patterns.

Results

Static Patterns

We excluded participants whose discrimination accuracy was not above 75% because we needed to estimate a discrimination threshold that corresponded to 75% discrimination accuracy. Hence, seven participants were excluded, and data from 14 participants were used for further analysis. We also used a more tolerant criterion and only excluded participants whose overall accuracy was not above 70%. This criterion only excluded four participants, and data from 17 participants were included for further analysis. This more tolerant criterion led to very similar results.

Psychometric Functions of Pooled Responses

As illustrated in Figure 5.4, the results of pooled responses showed that concentric patterns were easiest to perceive in noise: 13.5% signal strength was enough for participants to achieve 75% discrimination accuracy. Radial, radial-part, and concentric-part patterns were slightly more difficult to perceive: 20% signal strength was required for people to achieve 75% discrimination accuracy. Translational patterns were the most difficult to perceive: they required more than 30% signal strength for 75% discrimination accuracy. In addition, the asymptotic discrimination accuracy for translational patterns was only around 90% even when signal strength approximated 100%. This asymptotic accuracy was lower than that for other patterns.

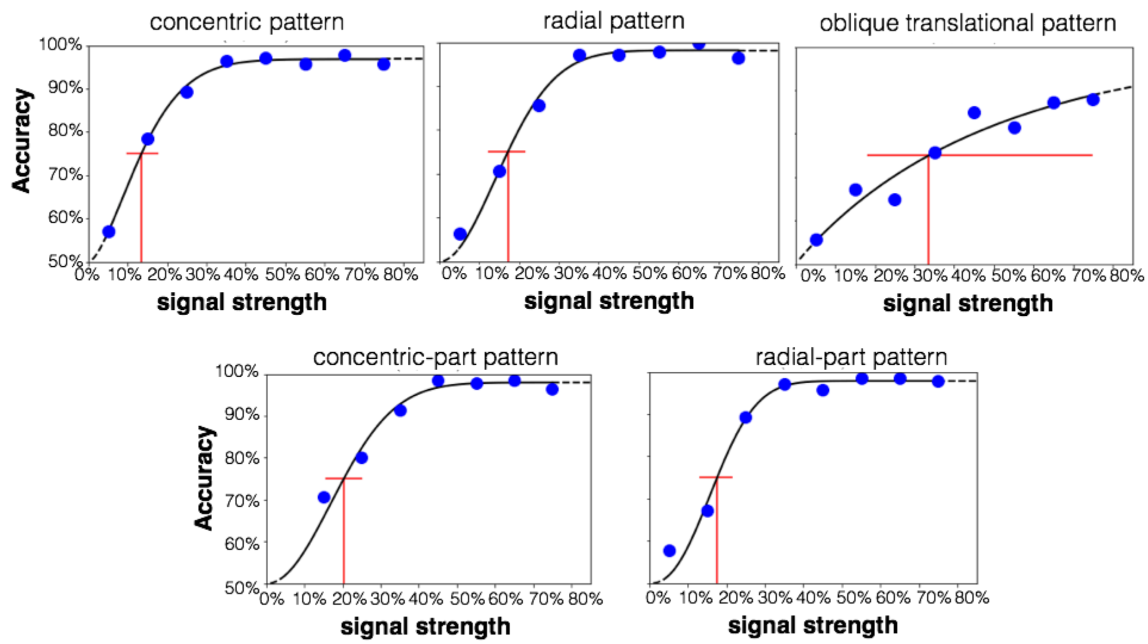


Figure 5.4 Psychometric functions of discrimination accuracies of five Glass patterns

Geometric Pattern	Discrimination Threshold	95% Confidence Interval of the threshold
Concentric Pattern	13.5%	[8.9%, 18.8%]
Radial Pattern	17.2%	[11.1%, 23.2%]
Translational Pattern	33.6%	[16.3%, 89.3%]
Concentric-part Pattern	20.3%	[14.4%, 26.7%]
Radial-part Pattern	17.6%	[12.0%, 22.5%]

Table 5.1 Psychometric function fitting results of pooled response data in Experiment 5.1: static patterns

Individual Fittings and Hypothesis Testing

For individual fitting results, we used repeated measures ANOVA to test whether discrimination thresholds were the same for different patterns. The results showed a significant difference in discrimination threshold ($F(4, 52) = 9.07$, $p < 0.0001$) among different geometric patterns. Post-hoc analyses showed that discriminating translational patterns required a significantly higher discrimination threshold than for the other four patterns ($p < 0.05$ in Tukey HSD testing). There were no significant differences in pairwise comparisons between these four patterns.

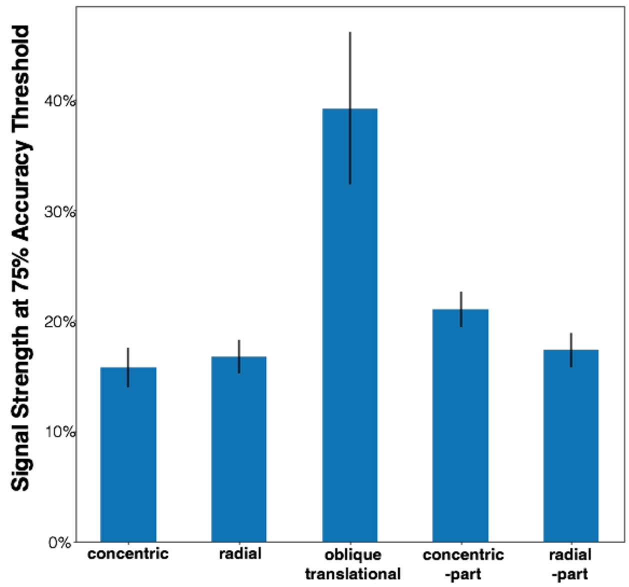


Figure 5.5 Discrimination thresholds of five Glass patterns

Motion Pattern

The overall accuracy of each participant in the motion pattern task was much higher than in the static pattern task, possibly because motion provided stronger signals for pattern perception. Thus, we had to exclude fewer participants from the analysis. Similar to the analysis of the static pattern task, we excluded participants whose accuracy was not above 75%. Two participants were excluded, and data from 19 participants were used for further analysis.

Psychometric Functions of Pooled Responses

From the pooled analysis, we found that people could perceive moving geometric patterns at a very low signal strength level, much lower than for static geometric patterns. As illustrated in Figure 5.6, the results showed that concentric and concentric-part patterns were easiest to perceive among noise in a motion stimulus. These two patterns had the lowest discrimination thresholds: participants could achieve 75% discrimination accuracy when signal strength was at about 10%. The other three patterns had higher discrimination thresholds, corresponding to 15% signal strength. The asymptotic detection accuracies were similar for all patterns.

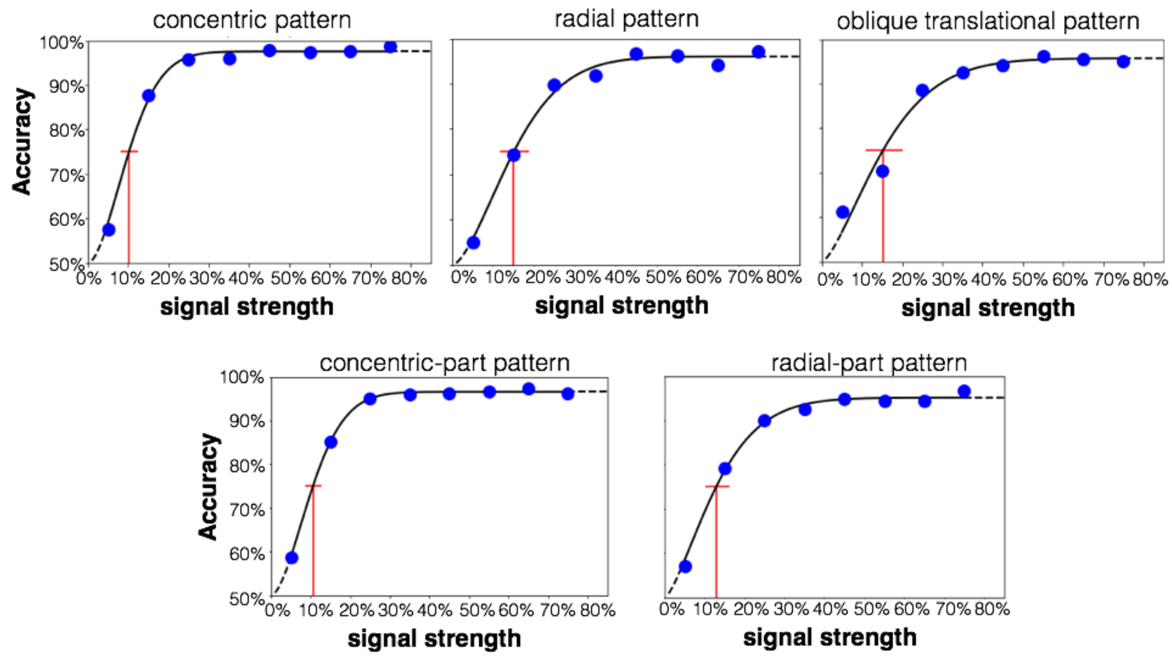


Figure 5.6 Psychometric functions of detection accuracies of five motion patterns

Geometric Pattern	Threshold at 75% Detection Accuracy	95% Confidence Interval of the threshold
Concentric Pattern	10.2%	[7.7%, 13.2%]
Radial Pattern	15.1%	[10.8%, 20.3%]
Translational Pattern	15.2%	[10.0%, 21.5%]
Concentric-part Pattern	10.6%	[8.2%, 13.4%]
Radial-part Pattern	12.9%	[9.5%, 17.2%]

Table 5.2 Psychometric function fitting results of pooled response data in Experiment 5.1: motion patterns

Individual Fittings and Hypothesis Testing

For individual fittings, we used repeated measures ANOVA to test whether discrimination thresholds were the same for different patterns. The results showed a significant difference in discrimination thresholds ($F(4, 72) = 5.74$, $p = 0.0005$) for different geometric patterns. Post-hoc analyses showed that the concentric patterns and concentric-part patterns had significantly lower thresholds than the translational patterns ($p < 0.05$ in Tukey HSD testing).

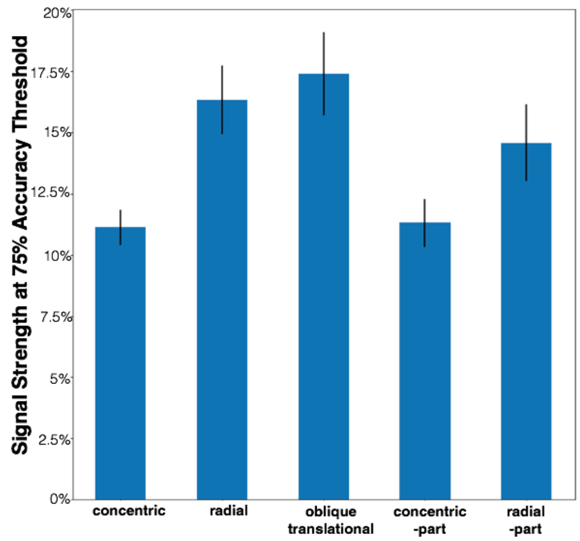


Figure 5.7 Discrimination thresholds of five motion patterns

Discussion

We asked participants to perceive five geometric patterns among perceptual noise, and found that there are indeed differences in perceiving different geometric patterns. In both static and motion stimuli, people can easily perceive concentric patterns, both intact or in part, among noise. Perception of these patterns only requires small amounts of signal that are aligned in correct spatial correspondences. People can also readily perceive a radial pattern among noise, although not as easily as a concentric one. In contrast, people cannot easily perceive a translational pattern. Perception of a translational pattern requires large amounts of signal that are aligned in correct correspondences. This result is counter-intuitive. The local parts of a translational pattern are similar to its global appearance; thus, perceiving such a pattern could be quick and easy in a Cartesian representation. In contrast, a concentric pattern requires integrating information distributed in space; a small segment of local information cannot reveal its global appearance. Thus, perceiving a concentric pattern should be slower if integration computation is required to connect local segments to its global circle shape. However, if a polar coordinate system is used, perceiving concentric patterns can be quick and easy: in a polar description of a concentric pattern, local parts are similar to its global appearance (Figure 5.2).

Why are there differences in the perceptibility of different geometric patterns? Wilson, Wilkinson, and Asaad (1997) showed similar low detection thresholds for concentric and radial patterns and attributed the differences in recognition sensitivity to a template matching process in higher-level visual areas, such as V4. They suggested that a global pattern template in these areas facilitates recognizing global patterns such as concentric and radial patterns, but does not help to recognize translational patterns (Wilson, Wilkinson, and Asaad, 1997). However, Dakin and Bex (2002) argued against this hypothesis, instead attributing a low detection threshold to the figure aperture in which the Glass patterns were displayed in the experiments conducted by Wilson, Wilkinson, and Asaad (1997). Namely, in their experiments, the Glass patterns were presented in a circular aperture, instead of a square aperture as in our experiments. However, Dakin and Bex's (2002) argument does not hold for Wilson's replication (Wilson & Wilkinson, 2003) or for our experiments. Explanations from Wilson and his colleagues do not apply to our experiment either. A template matching process in area V4 cannot explain similar advantages in perceiving moving circular patterns, because this area is traditionally treated as a part of the ventral pathway that identifies the properties of static objects (Desimone & Schein, 1987). Moving patterns, on the other hand, usually activate areas MT and MST (Celebrini & Newsome, 1994; Newsome & Pare, 1988) when random dot stimuli are used. Recent studies on motion streaks have shown that Glass patterns can also activate the area MST (Krekelberg, Dannenberg, Hoffmann, Bremmer, & Ross, 2003; Kourtzi & Kanwisher, 2000), rendering the V4 template matching explanation less convincing. Moreover, we also showed that people's perception of partial patterns (concentric-partial and radial-partial) is not significantly different from their perception of those intact patterns. To accommodate this result, a template matching explanation needs to either add a new template in V4 or describe an algorithm that quickly and easily completes a partial pattern into a full pattern. These additional requirements make the explanation of V4 template matching less likely to be the main reason for the easy perception of circular and radial patterns in our experiments.

We argue that the general perceptual advantages for circular and radial patterns should be attributed to the more fundamental process of computing local spatial correspondences. Specifically, the format of position representation should play a role in determining such perceptual advantages. We argued earlier that perceptual difficulties should be correlated with computational complexity in perception. Thus, our results suggest that concentric and radial patterns have simple descriptions in our visual system, and perceiving these patterns induces lower computational complexity than perceiving translational patterns. A Cartesian coordinate system should support simple models for translational patterns and lead to advantages in perceiving these translational patterns. However, we did not observe such advantages. Instead, we found that patterns with simple models in a polar system were always easier to perceive, in both static and motion conditions. Therefore, our results further indicate that the format of our visual spatial representation is polar.

Experiment 5.2 Mixed Pattern Detection

In Experiment 5.2, we directly tested perceptibility differences between two patterns. We mixed two patterns together into an ambiguous visual stimulus and asked participants to report what they perceived in that stimulus.

In each trial, we mixed two Glass patterns. For example, we arranged 40% of random dots into a spatial correspondence that gave rise to a concentric pattern, and 60% of them into a correspondence that gave rise to a translational pattern. Our visual system, however, does not know the ground truth of such an arrangement. Instead, it still needs to search for a reasonable spatial correspondence for each dot before fitting a model of final pattern perception. To perceive the concentric component, our visual system needs to fit a model of concentric patterns; and to perceive the translational component, our visual system needs to fit a model of translation patterns. During this competing model fitting process, our visual system may fit a model that has a simple description faster or more robustly, thus we should perceive the corresponding pattern more easily. As we discussed previously, whether a description is simple or complicated depends on the format of visual spatial representation. Therefore, the perceptual results of mixed patterns can provide further evidence concerning the format of our visual spatial representation.

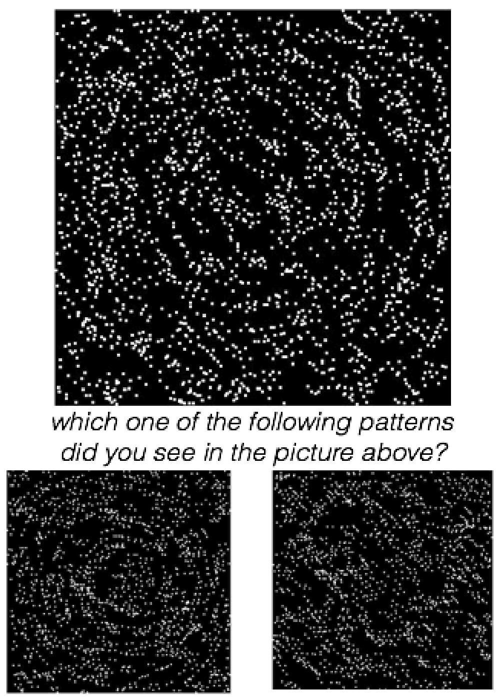


Figure 5.8 A schematic description of the discrimination task in Experiment 5.2

Methods

Participants

Another group of 20 Johns Hopkins University undergraduate and graduate students participated in the experiment in exchange for course credits. All participants reported normal or corrected-to-normal visual acuity and gave written informed consent. The experiment was approved by the Johns Hopkins University institutional review board (IRB).

Stimulus

Basic Component Pattern: Similar to Experiment 5.1, we designed five types of patterns as basic components to construct mixed displays. In addition to the ***concentric*** pattern, the ***radial*** pattern, and the ***oblique translational*** pattern, which were used in the Experiment 5.1, two new patterns were created: the fourth pattern, the ***horizontal translational*** pattern, had an appearance of multiple parallel horizontal lines; and the fifth pattern, the ***vertical translational*** pattern, had an appearance of multiple parallel vertical lines.

Mixed Pattern: In each stimulus, two of the aforementioned components were combined to construct a mixed pattern. Five mixed pairs were constructed: 1) concentric vs. oblique translational mixture, 2) concentric vs. radial mixture, 3) radial vs. oblique translational mixture, 4) oblique translational vs. horizontal translational mixture, and 5) oblique translational vs. vertical translational mixture. For each mixed pair, the proportion of one pattern varied from 10% to 90%, with a step size of 10%; accordingly, the proportion of the other pattern varied from 90% to 10%.

Design and Procedure

Each trial began with a 500ms fixation point. Subsequently, a mixture-pattern figure was briefly presented on the screen for 150ms. Later, two figures were presented side by side as candidates.

These two candidates were the two component patterns in the first figure. Participants needed to pick one based on what pattern they perceived in the first figure. A trial ended when the response was recorded.

Analysis

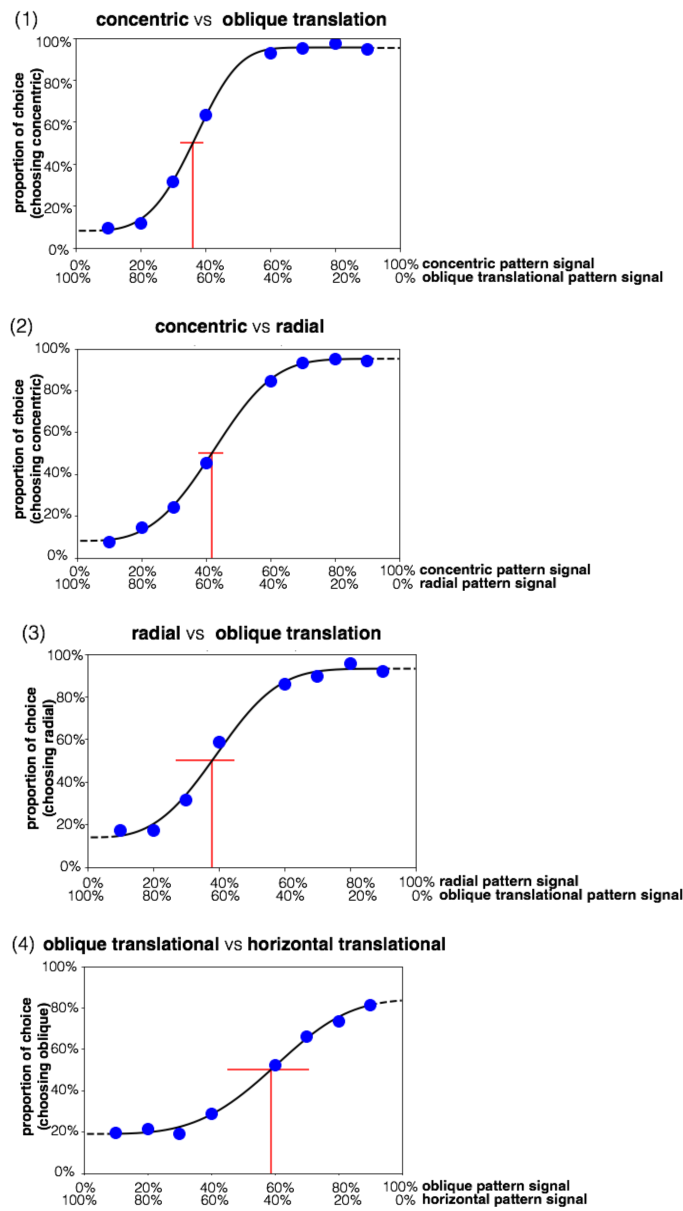
As in Experiment 5.1, we used the python package *psignifit* 4.0 (Schütt et al, 2016) and chose the Weibull function as the model to fit psychometric functions. For each mixture, we calculated the percentage of choosing one of the components in the mixture. This percentage increased as the signal proportion of the component increased from 10% to 90%. We fitted a psychometric function to describe the relationship between the percentage of choices and the signal proportion of the corresponding component. We used the maximum a posterior (MAP) estimation as the point estimation of an **equal-perceptibility threshold**: a signal proportion level of the component at which participants reported perceiving this component in 50% of trials, and perceiving the other component in 50% of trials as well. For example, in a concentric-radial mixture, with the signal proportion of a concentric pattern varying from 10% to 90%, the larger the signal proportion was, the more likely people were to perceive a concentric pattern in that mixture. An equal-perceptibility threshold was the signal proportion level of concentric patterns at which participants perceived a concentric component in the mixture in 50% of trials and a radial component in the other 50% of trials. We further extracted a 95% confidence interval of the threshold based on the threshold's posterior distribution.

The equal-perceptibility threshold reflected the relative difficulty in perception between two components in a mixture. If these two components were equally easy to perceive, the threshold should correspond to a 50%-50% mixture of the two components in which both patterns had the same signal strength. However, if one component was easier to perceive than the other, people still perceived that pattern instead of the other, more difficult one, even if that easy pattern possessed less than 50% of the overall signals. In sum, if an equal-perceptibility threshold corresponded to a mixture in which one pattern had a signal proportion of less than 50%, this pattern was regarded as “easier to perceive” than the other pattern in the mixture.

Results

First, we pooled response data from all participants together and fitted a single psychometric function for each mixture. This analysis resulted in five psychometric functions that described overall choices in each mixture. Figure 5.9 illustrates these fitted psychometric functions.

When a concentric pattern was mixed with a radial or a translational pattern (Figure 5.9 (1), (2)), people could easily perceive the concentric pattern in the mixture, even though it only possessed about 40% of the overall signal. Another way to look at the psychometric function is the following: when both components in a mixture had the same signal strength – i.e. both possessed 50% of the overall signal – people were more likely to perceive the concentric component in the mixture. These two understandings of psychometric functions converge to a single interpretation: perceiving concentric patterns was easier than perceiving radial or translational patterns. Similarly, the results of the psychometric function of the radial vs. oblique translational mixture (Figure 5.9 (3)) suggest that perceiving radial patterns was easier than perceiving oblique translational patterns. When two different translational patterns were mixed together, perceiving one pattern was not clearly easier or more difficult than perceiving the other. Furthermore, Figure 5.9 (4) and Figure 5.9 (5) show that a horizontal and a vertical pattern were easier to perceive than an oblique translational pattern, but the equal-perceptibility threshold was close to 50% of the signal strength of both components in a mixed pair, and the 95% confidence interval for each threshold included the 50%-50% signal strength of both components. These results indicate that these two types of translational patterns had no clear differences in perceptibility.



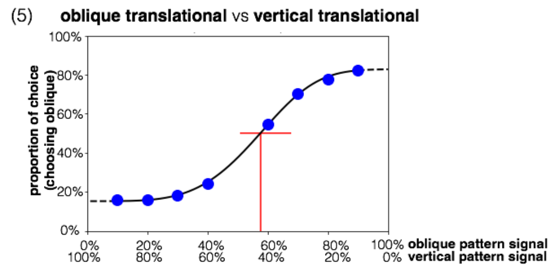


Figure 5.9 Six psychometric functions of detection accuracies of five mixed patterns

Geometric Pattern (Pattern A vs. Pattern B)	**Threshold at 50% Choice of Pattern A (Signal Strength of Pattern A)**	**95% Confidence Interval of the threshold (Signal Strength of Pattern A)**
Concentric (A) vs. Oblique Translational (B)	38.0%	[27.0%, 45.0%]
Concentric (A) vs. Radial (B)	41.9%	[37.7%, 45.4%]
Radial (A) vs. Oblique Translational (B)	36.2%	[32.2%, 39.6%]
Oblique Translational (A) vs. Horizontal Translational (B)	58.9%	[45.1%, 70.8%]
Oblique Translational (A) vs. Vertical Translational (B)	57.5%	[50.7%, 67.8%]

Table 5.3 Psychometric function fitting results of pooled response data in Experiment 5.2

In sum, the results of the psychometric functions showed that different geometric patterns were not equally easy to perceive: 1) concentric patterns were the easiest of them all; 2) concentric and radial patterns were easier than translational patterns; and 3) different translational patterns were similar in perceptibility, although horizontal and vertical translational patterns were slightly easier than oblique translational patterns.

Individual Fittings and Hypothesis Testing

Our analysis of individual fittings aimed to determine whether, for each mixed pair, there was a statistically significant difference in perception between two components in the mixture. We fitted a psychometric function to responses to each mixture for each participant. We extracted the equal-perceptibility threshold for each pair, and conducted five one-sample *t-tests* to answer our question. The null hypothesis was that the equal-perceptibility threshold corresponded to a signal strength of a 50% proportion of overall signals for each component (threshold = 50%). To correct for multiple comparisons in the hypothesis testing, we used a Bonferroni corrected alpha value $\alpha = 0.01$ ($\alpha = 0.05$ / 5 comparisons) as the critical value in hypothesis testing. We also used a stricter criterion of $\alpha = 0.002$ ($\alpha = 0.01$ / 5 comparisons) as the critical value to reject the null hypothesis. One-sample *t-tests* showed three significant results where the equal-perceptibility threshold was significantly different from 50%: (1) concentric vs. oblique translational mixture ($t(19) = -11.99$, $p = 2.61 \cdot 10^{-10}$), (2) concentric vs. radial mixture ($t(19) = -4.38$, $p = 3.25 \cdot 10^{-4}$), and (3) radial vs. oblique translational mixture ($t(19) = -7.71$, $p = 2.88 \cdot 10^{-7}$). On the other hand, two tests failed to reach a significant difference: (1) oblique translational vs. horizontal translational mixture ($t(19) = 1.19$, $p = 0.248$) and (2) oblique translational vs. vertical translational mixture ($t(19) = 2.01$, $p = 0.0586$).

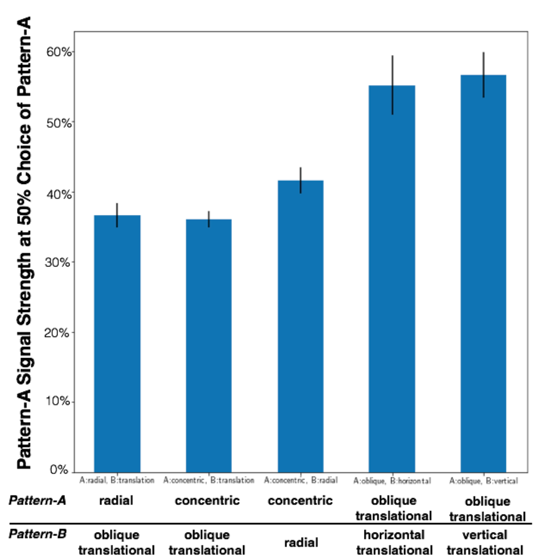


Figure 5.10 Signal strength at 50% choice of one pattern in each mixed pair

The results from the individual fittings and hypothesis testing were the same as for the previous psychometric function fitting to pooled data: perceiving concentric patterns was easier than perceiving radial and translational patterns, and perceiving radial patterns was easier than perceiving translational patterns. When different translational patterns were mixed together, no significant differences were found in perception between the two patterns.

Overall, our results show that concentric and radial patterns are perceived more easily in visual perception than translational patterns are. We reason that concentric and radial patterns have simple descriptions in a polar coordinate system, whereas translational patterns have simple descriptions in a Cartesian coordinate system. Our findings of easy perception of concentric and radial patterns suggest that our visual system uses a polar coordinate system to compute spatial correspondence and integrate global geometric pattern perception with these polar descriptions. Accordingly, the visual system can easily perceive geometric patterns that have simple descriptions in a polar coordinate system.

Discussion

When two patterns are mixed in one visual stimulus, two perceptual integration processes must compete with each other to arrive at a stable perception of that ambiguous stimulus. Each process is responsible for extracting signals of a geometric pattern. We hypothesize that this signal extraction computation requires fitting a model that describes geometric patterns. We reason that when two perceptual integration processes compete, one process that fits a simpler model will win the competition and dominate the visual perception of the mixed stimulus. Thus, people perceive the pattern of a simpler model between the two components. In this experiment, we asked participants to report their perception of ambiguous stimuli that were made by different geometric patterns including concentric circle patterns, radial line patterns, and translational line patterns. By investigating the stable perception of each ambiguous stimulus, we examined which geometric patterns have models with simple descriptions, and which have complicated descriptions.

Our results show that perceiving concentric patterns is significantly easier than perceiving radial and translational patterns, and perceiving radial patterns is significantly easier than perceiving translational patterns. Compared with translational patterns, both concentric and radial patterns have simpler descriptions in a polar coordinate system, but more complicated descriptions in a Cartesian coordinate system. Thus, easily perceiving concentric and radial patterns suggests that our visual spatial representation uses a polar-like coordinate system to process locations in space and compute spatial correspondences to integrate local information into a global pattern. These results are consistent with the results of Experiment 5.1, and provide further evidence to support a polar format in visual spatial representation.

Furthermore, we found that different translational patterns yielded no significant differences in perception. When two translational patterns were mixed, people did not report one pattern to be significantly easier to perceive than the other. This is surprising, because a clear oblique effect exists in stimulus orientation perception: people have superior perception, such as more precise orientation discrimination ability, of horizontal and vertical orientations than of oblique orientations (Appelle, 1972). One translational pattern is essentially a set of multiple small oriented segments, so it could be that horizontal and vertical translational patterns are easier to perceive. However, our results do not show such an oblique effect. Both pooled fitting results and individual fitting results indicate that vertical and horizontal translational patterns are only slightly easier to perceive than oblique translational patterns, though none of these perceptual advantages reached statistical significance. It is possible that with a polar system, different translational patterns are equally difficult to describe and compute and local orientation differences in these patterns do not contribute to the main computational bottleneck, making different translational patterns equally difficult to perceive.

Chapter 6 General Discussion

In this thesis, we aimed to understand how a human mind represents positions in space. We reasoned that an understanding of position representations in the mind should address two important questions (McCloskey, 2009). First, what information is represented in the mind? We need to identify the correct *contents* that the mind tries to describe and extract from the physical world. Second, how is such information represented in the mind? We need to clarify the *format* that the mind uses to organize information within its system. The work presented in this thesis is an attempt to answer these two questions. Our empirical studies have provided preliminary answers from multiple aspects in the visual domain concerning visual perception, visually guided actions, and eye movements.

Our research started with a case study of a neurological patient, MDK, who is diagnosed with posterior cortical atrophy. MDK provided us with unique opportunities to observe rich and diverse contents of position representations in his mind: positions are described in different reference frames, and these descriptions co-exist and interact with each other depending on the context. In chapter 2, we presented MDK's pointing performance when different stimuli and requirements were introduced to him. Starting from a simple observation that MDK had a systematic deficit when he was pointing to and looking at peripheral positions, we found that his visually guided actions were planned in a space that was centered on his fixation. This reference frame of position representation was maintained by his online visual feedback. In addition to this egocentric representation of space, he also had position representations that depended on objects in space. He engaged such object-centered representations in tasks where he needed to interact with the whole object, such as pointing to both ends of an oriented bar. In addition to investigating the reference frame of position representations, we also probed the format of

MDK's position representations. We observed that his pointing responses, though always biased towards the fixation, maintained their direction more or less accurately. A similar phenomenon occurred when he saccaded to peripheral positions. These results suggest that his mind always separately represents a peripheral position's angular bearing and distance, similar to a polar coordinate system. However, based on a close examination of his eye movement trajectories, we found that a polar coordinate system cannot explain his saccade execution process. The horizontal and vertical subcomponents of his oblique saccades were rarely synchronized. Therefore, we argue that the eye movement system separately encodes horizontal and vertical translation information, similar to a Cartesian coordinate system.

Inspired by our testing on MDK, we concentrated in our next three chapters on revealing the format of position representations. Specifically, we tried to understand when and what coordinate system is used in the visual domain to represent positions. A format may elude direct measurement, but it can be revealed by subtle clues with careful examination. We tried to unveil the underlying formats by 1) analyzing variances and biases of behavioral responses, 2) identifying primitives in perceptual computations, and 3) comparing complexities of recognition algorithms.

In Chapter 3, we analyzed variances and biases in people's responses by asking them to repeatedly respond to a single dot in their peripheral visual field. These simple tasks yielded abundant response data. Using these data, we fitted probabilistic models that described distributions of positional variances in two different formats: a Cartesian coordinate format and a polar coordinate format. Our model comparison results showed that a polar model can better capture the behavioral variances in visual perception and visually guided actions, whereas a Cartesian model can better capture the variances in eye movements. Biases in responses also met our model predictions. Biases were clear when we were able to accumulate them with an iterative process. Similar to previous results, a polar model could explain the accumulative biases in visually guided pointing better than a Cartesian model could. Together, the results from modeling variance and bias suggest that polar coordinate systems and Cartesian coordinate systems are used to represent positions in different visual subsystems.

In Chapter 4, we identified primitive information in a position representation by adapting classic Vernier discrimination tasks. Primitive information is information that is native in a representation and can be accessed easily without further transformation. We first showed that angular discrimination was relevant to the reference angle of a position but irrelevant to the distance of this position from the fixation. Similarly, distance judgment was also accurate when positions had different angular bearings. These results indicate that angle and distance may be encoded in primitives in the visual representation of positions. However, there was a difference between judgment performances when stimuli appeared on cardinal axes versus on oblique axes. We posited that this difference may be due to the use of different heuristics to judge distances in these two conditions. Furthermore, we showed that people's subjective perception of a straight line had systematic biases towards the circumference of a circle centered at fixation. A similar performance difference was observed between cardinal and oblique conditions, but the perceptual bias prevailed in all straight line alignment performances. Together, our results suggest that distance and angle, rather than horizontal and vertical translations, are encoded as primitives in the position representation in perception, as a polar coordinate system. However, people can use heuristic strategies when they judge vertical and horizontal alignments. The same heuristics can help decrease perceptual biases in aligning positions on straight lines on the one hand, and can induce perceptual biases in judging distances of positions on the other hand.

In Chapter 5, we examined the complexity of possible algorithms in vision. As we argued previously, the complexity of an algorithm depends on representations. We chose to study pattern recognition algorithms that integrate local information to form a global perception. We tested people's perception of global patterns comprising scattered local dots. We also added random dots as noise to test how perceptual pattern recognition algorithms were robust to noisy input. We found that people could easily detect concentric circle and radial line patterns among noise, but not translational line patterns. In addition, when concentric or radial patterns were intermixed with translational patterns, people first recognized concentric or radial patterns rather than translational patterns. These results suggest that algorithms for perceiving a concentric pattern and radial line have lower complexity than those for perceiving translational lines. The underlying reason for this may be a polar-like representation of positions in the visual perception system. In a polar coordinate system, concentric and radial patterns have simple, linear descriptions, and only a simple algorithm is needed to describe these patterns. This simple algorithm can easily extract the information it needs to recognize concentric or radial patterns among noise, thus leading to a low pattern detection threshold. Translational patterns, in contrast, have complicated descriptions in a polar coordinate system. Their recognition relies on translational information, which has a simple, linear description in a Cartesian coordinate system. To retrieve translational information in a polar coordinate system, where such information is usually hidden in the representation, algorithms need to use computationally expensive processing steps. In turn, this complex algorithm requires more signals to recover translational patterns, thus leading to a high pattern detection threshold.

In sum, we have provided empirical evidence to show that visual perception and visually guided actions use a polar coordinate system to represent positions, whereas the eye movement systems use a Cartesian one. Together with our observations of the multiple reference frames used by MDK, we have presented our characterization of position representations in the visual domain, including both their reference frames and formats.

Our characterization may not seem comprehensive because we only tested two alternative hypotheses. However, it should be a good approximation. First, MDK's dissociation of deficits in angle and distance information suggests that these two pieces of information should have separate abstraction and manipulation. Second, previous and current studies can capture variances in visual behaviors very well using either a polar model or a Cartesian model (Greenwood, Szinte, Sayim, & Cavanagh, 2017). Third, experimental and theoretical neuroscience research also provides supporting evidence for a polar coordinate system in visual perception. Retinotopic mapping studies and mathematical formulations have indicated that information in early visual cortices undergoes a logarithmic-polar coordinate transformation from information on the retina. Such a transformation may lead to a natural description of retinal images in a polar space (Schwartz, 1980a; 1980b). Furthermore, electrophysiological recording studies have also found neurons in the parietal cortex that respond to angles and distances of a position in three-dimensional space within a body-centered reference frame (Guenther, Bullock, Greve, & Grossberg, 1994).

A detailed characterization of representation is important because this characterization can inform us on how our mind works. Marr (1982) extensively discussed the role of representation in a computational theory of mind. A representation is a system that can describe a given entity by using symbols or combinations of symbols (Marr, 1982). It not only determines a symbolic form of inputs and outputs of a computational process, but also affects the nature of input-output transformation algorithms, such as their complexity and constraints. Our minds work on information conveyed by a representation (or equally, a representing system) to make sense of the represented entity, including elements and rules of that entity.

One aspect of understanding a representing system, or a representation, in the mind is actually to identify the represented entity. This aspect corresponds to our first endeavor to understand what reference frames are used in representations. Even after identifying a computational problem, as emphasized by Marr, we may still not know what to represent in that problem. For example, even after we specify a problem as calculating the position of a part of an object so that we can reach it, we still need to determine whether the position is described in a body-centered reference frame or an object-centered reference frame. MDK's testing showed such a complication in his unimanual versus bimanual pointing responses. This complication is not obvious in Marr's work, because he often unambiguously identified the represented entity in his discussion, such as numbers, symphonies, words, or shapes. In open-ended research on representations in the mind, however, what is represented by the mind is unknown. Luckily, modern psychological and neuroscience research has spent a great deal of effort mainly trying to identify the represented entity, at least for spatial perception (Golomb, & Kanwisher, 2011; Guenther, Bullock, Greve, & Grossberg, 1994).

The other aspect of understanding a representing system, which was emphasized in Marr's original discussion but has sometimes been ignored in recent experimental research, is to dissect that system and examine its own elements and rules. This aspect corresponds to our second endeavor to unveil the format of representations. This knowledge informs us on how the representing system organizes information that it needs to describe and support computational algorithms. Thus, understanding the format of representations can guide us in seeking knowledge at the algorithmic level of a computational theory of mind. As a result, an improved understanding of algorithms may be manifested as constructive, step-by-step solutions, rather than existential proofs of such algorithms or box-arrow diagrams. Future research may create or examine multiple possible algorithms, possibly adopting ones from computer vision research, and design experiments to determine which one is more likely to be used by the human mind.

To understand these two aspects of a representation in a mind is difficult. We need to utilize different approaches to gain insight into what a representation may be and what symbols and rules it uses. First, a physical realization of a cognitive system may shed light on the representation. Neuroscience research may provide clues from neuron activities in response to types of information, and guide us to abstract variables that should exist in a representing system. For example, the combination of grid-cell systems and place cell systems is promising to help our understanding of representations involved in navigation. However, this approach faces difficulties in bridging the gaps between levels of analysis. Second, machine learning algorithms may automatically learn proper representations for us. Recent learning algorithms do not constrain themselves in learning mapping functions between input data and output data; many of them are designed to learn a good representation of data so that further processing can easily separate information and effectively transform useful pieces to solve problems (Bengio, Courville, and Vincent, 2016). This approach, however, faces difficulties in interpreting and understanding the learned representations. Third, independent advances in mathematics, sciences, and technologies, especially those concerning how to build machines, may unexpectedly provide a solution to understand representations in our mind. Designing a system from scratch can force us to think about requirements and constraints for our design. Thus, knowing each piece of the system enables us to articulate algorithmic solutions. New representations of data may emerge in the process as we try to meet needs or overcome constraints when we determine features in design or implement a design idea. These new representations may already be used in our mind to describe and process information as a result of natural evolution. History has shown that the better we can model the world, the better we may reverse-engineer our mind.

Reference

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: the" oblique effect" in man and animals. Psychological bulletin, 78(4), 266.

Afraz, S. R., & Cavanagh, P. (2008). Retinotopy of the face aftereffect. Vision research, 48(1), 42-54.

Afraz, A., & Cavanagh, P. (2009). The gender-specific face aftereffect is based in retinotopic not spatiotopic coordinates across several natural image transformations. Journal of Vision, 9(10), 10-10.

Andersen, R. A. (1995). Encoding of intention and spatial location in the posterior parietal cortex. Cerebral Cortex, 5(5), 457-469.

Bahill, A. T., & Stark, L. (1975). Neurological control of horizontal and vertical components of oblique saccadic eye movements. Mathematical Biosciences, 27(3), 287-298.

Bahill, A. T., & Stark, L. (1977). Oblique saccadic eye movements: Independence of horizontal and vertical Stark, L. (1979). The trajectories of saccadic eye movements. Scientific American, 240(1),channels. Archives of Ophthalmology, 95(7), 1258-1261.

Bender, D. B. (1981). Retinotopic organization of macaque pulvinar. Journal of Neurophysiology, 46(3), 672-693.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35(8), 1798-1828.

Benson, D. F., & Greenberg, J. P. (1969). Visual form agnosia: A specific defect in visual discrimination. Archives of Neurology, 20(1), 82-89.

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nature neuroscience, 19(1), 150-157.

Brainard, D. H. (1997). The psychophysics toolbox. Spatial vision, 10, 433-436.

Breitmeyer, B. G., Kropfl, W., & Julesz, B. (1982). The existence and role of retinotopic and spatiotopic forms of visual persistence. *Acta psychologica*, *52*(3), 175-196.

Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., & Wiener, M. C. (2001).Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 356(1407), 299-330.

Caminiti, R., Johnson, P. B., Galli, C., Ferraina, S., & Burnod, Y. (1991). Making arm movements within different parts of space: the premotor and motor cortical representation of a coordinate system for reaching to visual targets. The Journal of neuroscience, 11(5), 1182-1197.

Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. Trends in cognitive sciences, 14(4), 147-153.

Celebrini, S., & Newsome, W. T. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. Journal of Neuroscience, 14(7), 4109-4124.

Corkin, S. (1984). Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in HM. In Seminars in neurology (Vol. 4, No. 2, pp. 249-259).

Dabaghian, Y., Brandt, V. L., & Frank, L. M. (2014). Reconceiving the hippocampal map as a topological template. Elife, 3, e03476.

REFERNCE

Dakin, S. C., & Bex, P. J. (2002). Summation of concentric orientation structure: seeing the Glass or the window?. Vision Research, 42(16), 2013-2020.

Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. Journal of neurophysiology, 57(3), 835-868.

Dessalegn, B., & Landau, B. (2008). More than meets the eye the role of language in binding and maintaining feature conjunctions. Psychological Science, 19(2), 189-195.

Driver, J., & Halligan, P. W. (1991). Can visual neglect operate in object-centred co-ordinates? An affirmative single-case study. Cognitive Neuropsychology, 8(6), 475-496.

Eagleman, D. M., & Sejnowski, T. J. (2007). Motion signals bias localization judgments: A unified explanation for the flash-lag, flash-drag, flash-jump, and Frohlich illusions. Journal of vision, 7(4), 3-3.

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cerebral cortex, 7(2), 181-192.

Ermentrout, G. B., & Cowan, J. D. (1979). A mathematical theory of visual hallucination patterns. Biological cybernetics, 34(3), 137-150.

Feldman, J. (2003). The simplicity principle in human concept learning. Current Directions in Psychological Science, 12(6), 227-232.

Fischer, J., Spotswood, N., & Whitney, D. (2011). The emergence of perceived position in the visual system. Journal of Cognitive Neuroscience, 23(1), 119-136.

Fischer, J., & Whitney, D. (2012). Attention gates visual coding in the human pulvinar. Nature communications, 3, 1051.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning. New York: Springer series in statistics.

Gallistel, C. R. (1990). The organization of learning. The MIT Press.

Gallistel, C. R., & Matzel, L. D. (2013). The neuroscience of learning: beyond the Hebbian synapse. Annual review of psychology, 64, 169-200.

Gazzaniga, M. S. (2004). The cognitive neurosciences. MIT press.

Glass, L. (1969). Moire effect from random dots. Nature, 223(5206), 578

Glass, L., & Switkes, E. (1976). Pattern recognition in humans: Correlations which cannot be perceived. Perception, 5(1), 67-72.

Golomb, J. D., & Kanwisher, N. (2011). Higher level visual cortex represents retinotopic, not spatiotopic, object location. Cerebral Cortex, bhr357.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. Nature, 349(6305), 154.

Greenwood, J. A., Szinte, M., Sayim, B., & Cavanagh, P. (2017). Variations in crowding, saccadic precision, and spatial localization reveal the shared topology of spatial vision. Proceedings of the National Academy of Sciences, 201615504.

Grossman, G. E., & Robinson, D. A. (1988). Ambivalence in modelling oblique saccades. Biological cybernetics, 58(1), 13-18.

Guenther, F. H., Bullock, D., Greve, D., & Grossberg, S. (1994). Neural representations for sensorimotor control. III. Learning a body-centered representation of a three-dimensional target position. Journal of cognitive neuroscience, 6(4), 341-358.

Halligan, P. W., Fink, G. R., Marshall, J. C., & Vallar, G. (2003). Spatial cognition: evidence from visual neglect. Trends in cognitive sciences, 7(3), 125-133.

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. Cognition, 61(3), 195-232.

REFERNCE

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language.Cognitive psychology, 39(1), 3-36.

Holmes, G. (1944). The organization of the visual cortex in man. Proceedings of the Royal Society of London. Series B, Biological Sciences, 132(869), 348-361.

Hubel, D. H., & Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. Proceedings of the Royal Society of London B: Biological Sciences, 198(1130), 1-59.

Kastner, S., DeSimone, K., Konen, C. S., Szczepanski, S. M., Weiner, K. S., & Schneider, K. A. (2007). Topographic maps in human frontal cortex revealed in memory-guided saccade and spatial working-memory tasks. Journal of neurophysiology, 97(5), 3494-3507.

Katyal, S., Zughni, S., Greene, C., & Ress, D. (2010). Topography of covert visual attention in human superior colliculus. Journal of neurophysiology, 104(6), 3074-3083.

King, W. M., Lisberger, S. G., & Fuchs, A. F. (1986). Oblique saccadic eye movements of primates. Journal of Neurophysiology, 56(3), 769-784.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In Spatial cognition (pp. 1-17). Springer Berlin Heidelberg.

Knapen, T., Rolfs, M., Wexler, M., & Cavanagh, P. (2009). The reference frame of the tilt aftereffect. Journal of Vision, 10(1), 8-8.

Knudsen, E. I. (1982). Auditory and visual maps of space in the optic tectum of the owl. The Journal of Neuroscience, 2(9), 1177-1194.

Knudsen, E. I., & Brainard, M. S. (1995). Creating a unified representation of visual and auditory space in the brain. Annual review of neuroscience, 18(1), 19-43.

Krekelberg, B., Dannenberg, S., Hoffmann, K. P., Bremmer, F., & Ross, J. (2003). Neural correlates of implied motion. Nature, 424(6949), 674.Prazdny, K. (1984). On the perception of Glass patterns. Perception, 13(4), 469-478.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. Journal of cognitive neuroscience, 12(1), 48-55.

Lacquaniti, F., Guigon, E., Bianchi, L., Ferraina, S., & Caminiti, R. (1995). Representing spatial information for limb movement: role of area 5 in the monkey. Cerebral cortex, 5(5), 391-409.

Landau, B., & Lakusta, L. (2009). Spatial representation across species: Geometry, language, and maps. Current opinion in neurobiology, 19(1), 12-19.

Landau, B., & Hoffman, J. E. (2012). Spatial representation: From gene to mind. Oxford University Press.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. Nature neuroscience, 4(5), 533-539.

Marr (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. New York: Freeman.

Mathôt, S., & Theeuwes, J. (2013). A reinvestigation of the reference frame of the tilt-adaptation aftereffect. Scientific reports, 3.

McCloskey, M. (2009). Visual reflections: A perceptual deficit and its implications. New York:Oxford University Press.

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). Journal of Neuroscience, 8(6), 2201-2211.

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat.Experimental neurology, 51(1), 78-109.

REFERNCE

O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J., & Maguire, E. A. (1998). Place cells, navigational accuracy, and the human hippocampus.Philosophical Transactions of the Royal Society of London B: Biological Sciences, 353(1373), 1333-1340.

O'Keefe, J., & Conway, D. H. (1978). Hippocampal place units in the freely moving rat: why they fire where they fire. Experimental Brain Research,31(4), 573-590.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain research, 34(1), 171-175.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial vision, 10(4), 437-442.

Perenin, M. T., & Vighetto, A. (1988). Optic ataxia: a specific disruption in visuomotor mechanisms. Brain, 111(3), 643-674.

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. Sociological Methodology, 25, 111-163.

Ross, J., Morrone, M. C., & Burr, D. C. (1997). Compression of visual space before saccades. Nature, 386(6625), 598-601.

Rossetti, Y., Pisella, L., & Vighetto, A. (2003). Optic ataxia revisited. Experimental Brain Research, 153(2), 171-179.

Schneider, K. A., & Kastner, S. (2005). Visual responses of the human superior colliculus: a high-resolution functional magnetic resonance imaging study. Journal of neurophysiology, 94(4), 2491-2503.

Schluppeck, D., Curtis, C. E., Glimcher, P. W., & Heeger, D. J. (2006). Sustained activity in topographic areas of human posterior parietal cortex during memory-guided saccades. The Journal of neuroscience, 26(19), 5098-5108.

Schluppeck, D., Glimcher, P., & Heeger, D. J. (2005). Topographic organization for delayed saccades in human posterior parietal cortex. Journal of neurophysiology, 94(2), 1372-1384.

Schwartz, E. L. (1980a). A quantitative model of the functional architecture of human striate cortex with application to visual illusion and cortical texture analysis. Biological Cybernetics, 37(2), 63-76.

Schwartz, E. L. (1980b). Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. Vision research, 20(8), 645-669.

Schwartz, E. L. (1983). Cortical mapping and perceptual invariance: a reply to Cavanagh. Vision Research, 23(8), 831-835.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, *5*(1), 3-55.

Sheth, B. R., & Shimojo, S. (2001). Compression of space in visual memory. Vision research, 41(3), 329-341.

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. Vision research, 122, 105-123.

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. Trends in cognitive sciences, 13(11), 488-495.

Sinico, M., Parovel, G., Casco, C., & Anstis, S. (2009). Perceived shrinkage of motion paths. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 948.

Soechting, J. F., Tillery, S. H., & Flanders, M. (1990). Transformation from head-to shoulder-centered representation of target direction in arm movements. Journal of Cognitive Neuroscience, 2(1), 32-43.

REFERNCE

Tolman, E. C. (1948). Cognitive maps in rats and men. Psychological review,55(4), 189.

Turi, M., & Burr, D. (2012). Spatiotopic perceptual maps in humans: evidence from motion adaptation. Proceedings of the Royal Society of London B: Biological Sciences, 279(1740), 3091-3097.

Tsal, Y., & Bareket, T. (1999). Effects of attention on localization of stimuli in the visual field. Psychonomic Bulletin & Review, 6(2), 292-296.

Tsal, Y., & Bareket, T. (2005). Localization judgments under various levels of attention. Psychonomic Bulletin & Review, 12(3), 559-566.

Wenderoth, P., & Wiese, M. (2008). Retinotopic encoding of the direction aftereffect. Vision research, 48(19), 1949-1954.

Wilson, H. R., Wilkinson, F., & Asaad, W. (1997). Concentric orientation summation in human form vision. Vision research, 37(17), 2325-2330.

Wilson, H. R., & Wilkinson, F. (2003). Further evidence for global orientation processing in circular Glass patterns. Vision research, 43(5), 563-564.

Wurtz, R. H., & Goldberg, M. E. (1972). Activity of superior colliculus in behaving monkey. III. Cells discharging before eye movements. J Neurophysiol, 35(4), 575-586.

Yantis, S. (2013). Sensation and perception. Palgrave Macmillan.

Curriculum Vitae

Feitong Yang was born in 1989 in Changsha, Hunan, China. He did his undergraduate work at Peking University where he received B.Sc. in Psychology. He worked with Dr. Sheng Li on perceptual decision-making. He enrolled the Psychological and Brain Sciences Ph.D. program at Johns Hopkins University in 2013 under the mentorship of Dr. Jonathan Flombaum. During the graduate research, he also had a long-term collaboration with Dr. Michael McCloskey in testing the neuropsychological patient, MDK. In the meantime, he enrolled in the Applied Mathematics and Statistics M.S.E. program at Johns Hopkins University in 2017. He served as a teaching assistant for five courses offered in the Department of Psychological and Brain Sciences and offered two university-wide intersession courses to undergraduates in 2015 and 2016.